Structural studies on proteins involved in angiogenesis

Mitsiki, Eirini

Award date:
2007

Awarding institution:
University of Bath

STRUCTURAL STUDIES ON PROTEINS INVOLVED IN ANGIOGENESIS

Submitted by Eirini Mitsiki

for the degree of Ph.D.

of the University of Bath

2nd October 2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Contents

List of figures

List of tables

Acknowledgements

I would like to thank my supervisor Prof. K. Ravi Acharya for giving me the opportunity to do this project as well as for his constant support and guidance. I would also like to thank my assessor Dr. Vasanta Subramanian for the constructive comments throughout my PhD.

I would also like to thank the University of Bath for the post-graduate research studentship. I am thankful to the station scientists at the Synchrotron Radiation Source, Daresbury Laboratory, UK for their support during X-ray data collection.

I would also like to thank my colleagues past/present in the laboratory Drs. Evangelia Chrysina, Anastassios Papageorgiou and Gayatri Chavali for the work they carried out before me on the hTP project. Special thanks go to Drs. Stephen Prior, Matthew Baker, Daniel Holloway, Shalini Iyer, Nethaji Thiyagarajan, Kenneth Holbourn and Miss Konstantina Kazakou and for their support and advice during my PhD.

Last but not least, I would like to thank my brother and my parents as well as relatives and friends for their support and understanding!

Abstract

Angiogenesis, the sprouting of new capillaries from preexisting microvasculature, is a key phenomenon for many physiological and pathological processes. Physiological angiogenesis occurs during embryogenesis and for a healthy adult it occurs during wound healing and the female ovarian/menstrual cycle. Uncontrolled or dysfunctional angiogenesis on the other hand, has been shown to be associated with ischemia of the heart, atherosclerosis, rheumatoid arthritis and tumour growth and metastasis. Control of angiogenesis occurs *via* a signal network of activators and repressors, known as angiogenesis-stimulating factors (angiogenic factors) and angiogenesis inhibitors (angiostatins). Proteins involved in this key process include nucleoside phosphorylases as well as the tissue inhibitors of metalloproteinases (TIMPs).

Nucleoside phosphorylases catalyse the reversible phosphorolysis of purine and pyrimidine nucleosides, a reaction of key importance for the nucleotide salvage pathway. Pyrimidine phosphorylases (PyNPs), which cleave the glycosidic bond of pyrimidines readily catalyses the reversible phosphorolysis (having primarily a catabolic function) of thymidine and 2'-deoxyuridine to their respective base and to 2-deoxy-D-ribose-1-phosphate, as well as that of some pyrimidine analogues. Thymidine phosphorylase (TP) was found to induce the [^{3}H] thymidine incorporation, endothelial cell migration *in vitro* and angiogenesis *in vivo*. This is achieved by retaining a constant and correct supply of deoxyribonucleoside triphosphates (dNTPs) for DNA repair and replication. TP has also been shown to play a crucial role in pathological angiogenesis; TP overexpression has been related to various carcinomas while TP deficiency has been related to certain clinical conditions. Here we report the crystal structures of hTP in two forms: unbound (native hTP) and bound with a small molecule inhibitor, 5IUR, which greatly resembles the chemotherapeutic agent currently used, 5FUR. Additionally, mutagenesis and enzymatic activity assays provide some novel information on the significance of certain active site residues and their role in the phosphorolytic reaction catalysed by hTP.

Extracellular matrix (ECM), which surrounds tissues and organs, has been shown to influence cell behaviour. Turnover of ECM components is therefore a crucial process for the control of cellular behaviour. It is essential in morphogenesis and tissue remodeling, which occur during embryonic development, wound healing and angiogenesis, while misregulation of ECM components turnover, has been linked to a series of pathological conditions. Enzymes involved in proteolytic systems that ECM turnover include the matrix-degrading proteinases known as Matrix Metalloproteinases (MMPs) or matrixins. The role of these molecules is highly controlled under physiological conditions, primarily by endogenous inhibitors, known as the tissue inhibitors of metalloproteinases (TIMPs). TIMPs have attracted a lot of interest due to their potential use in therapeutics. Understanding the structural features that are essential for TIMP potency against MMPs is thus of outmost importance. A bioinformatic analysis of an 'ancestral' TIMP form, TIMP from *Drosophila melanogaster* (*d*TIMP) allows the understanding of the evolutionary relationships that link TIMPs from different species, and raises questions about the significance of certain conserved TIMP features.

Abbreviations

2dR1P	2-deoxy-D-ribose-1-phosphate
5FU	5-Fluorouracil
5IUR	5-Iodouracil
7DX	7-Deazaxanthine
ADAM	A Disintegrin and A Metalloproteinase
ADAMTS	A Disintegrin and A Metalloproteinase with a Thrombospondin motif
BAC	Bacterial Artificial Chromosome
BCA	BicinChoninic Acid
BSA	Bovine Serum Albumin
BsPyNP	*Bacillus stearothermophillus* Pyrimidine nucleoside phosphorylase
CAM	Chick chorioallantoic membrane
CATH	Class, architecture, topology, homologous superfamily
CCD	Charged Couple Device
CuKa	Copper cathode
D.O.	Dissolved oxygen
DMSO	Dimethyl Sulfoxide
***d*TIMP**	Tissue Inhibitor of Metalloproteinase from *Drosophila melanogaster species*
ECM	Extracellular matrix
EcTP	*Escherichia coli* Thymidine Phosphorylase
EDTA	ethylenediamine tetraacetic acid
ESI – TOF	Electro-Spray Ionisation – Time Of Flight
FGF	Fibroblast Growth Factor
FRET	Fluorescence Resonance Energy Transfer
GCUA	Graphical Codon Usage Analyser
GeV	Gigaelectron volt
GLS	Gliostatin

GST	Glutathione S-transferase
HPLC	High Pressure Liquid Chromatography
***h*TIMP**	Tissue Inhibitor of Metalloproteinase from *Human species*
hTP	Human Thymidine Phosphorylase
IBD	Irritable Bowel Disease
IL-8	Interleukin-8
IPTG	Isopropyl-beta-D-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
LB media	Luria-Bertani media
LSQ	Least Square refinement
MAD	Multiple-wavelength Anomalous Dispersion
MALDI – TOF	Matrix Assisted Laser Disorpion/Ionisation – Time Of Flight
MBP	Maltose Binding Protein
MIR	Multiple Isomorphous Replacement
MIRAS	Multiple-wavelength Anomalous Dispersion and Anomalous Scattering
MLH	Maximum Likelihood refinement
MMP	Matrix Metalloproteinase
MNGIE	Mitochondrial neurogastrointestinal encephalomyopathy
MoKa	Molybdenum cathode
MPD	2-Methyl-2,4-pentanediol
MR	Molecular Replacement
***m*TIMP**	Tissue inhibitor of metalloproteinase from *Mus musculus* species
MT-MMP	Membrane-Type Matrix Metalloproteinase
NMR	Nuclear Magnetic Resonance
O.D.	Optical Density
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PD-ECGF	Platelet-Derived Endothelial Cell Growth Factor
PEG	Polyethylene glycol

PNP	Purine Nucleoside Phosphorylase
PVDF	Polyvinylidene Difluoride
PyNP	Pyrimidine Nucleoside Phosphorylase
RA	Rheumatoid Arthritis
RMSD	Root mean square deviation
***r*TIMP**	Tissue inhibitor of metalloproteinase from *Rattus norvegicus* species
SAD	Single-wavelength Anomalous Dispersion
SCOP	Structural classification of proteins
SDS-PAGE	Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis
SIR	Single Isomorphous Replacement
SIRAS	Single-wavelength Anomalous Dispersion and Anomalous Scattering
SRS	Synchrotron Radiation Source
TB media	Terrific Broth media
TCA	Trichloroacetic acid
TDR	Thymidine
TIMP	Tissue Inhibitor of Matrix Metalloproteinase
TNF-α	Tumour Necrosis Factor-α
TP	Thymidine Phosphorylase
TPI	5-chloro-6-[1-(2-iminopyrrolidinyl)methyl] uracil hydrochloride
UP	Uridine Phosphorylase
UV	Ultraviolet
VEGF	Vascular Endothelial Growth Factor
YAC	Yeast Artificial Chromosome

Chapter - I -

Introduction to X-ray crystallography

1.1 Introduction

Determination of the three dimensional structure of biological macromolecules using X-rays is known as X-ray crystallography. It is one of the most powerful techniques for structure determination allowing for details near or at atomic level resolution to be obtained. Information obtained from X-ray crystallography aids in the understanding of the structural characteristics, the enzymatic mechanism and protein – protein recognition of a biological macromolecule. Structure elucidation of enzymes interacting with substrates and identification of conformational changes or key amino acid residues can subsequently lead to the design and development of drugs that target the enzymatic activity. It is thus appropriate to say that X-ray crystallography attracts the interests of various scientific fields including biochemistry, physics, chemistry and mathematics. Advances in technology have lead to the development of synchrotrons (more powerful X-ray sources), fast computers, computer software and robotics that have greatly enhanced all the steps involved in the determination of macromolecular structures. The steps involved in macromolecular structure determination are shown in Figure 1.1.

Starting form the gene of interest, cloning, expression and purification lead to the production of the protein of interest. The purity of the protein is of paramount importance in crystallographic studies; obtaining good quality diffracting crystals can greatly enhance subsequent steps, such as data collection and processing. Significant improvements in the computer programs have greatly aided phase determination, model building and refinement, all of which are part of structure elucidation process. The significance of each of these steps in obtaining a macromolecular structure as well as understanding and correctly interpreting it, will be discussed in this section.

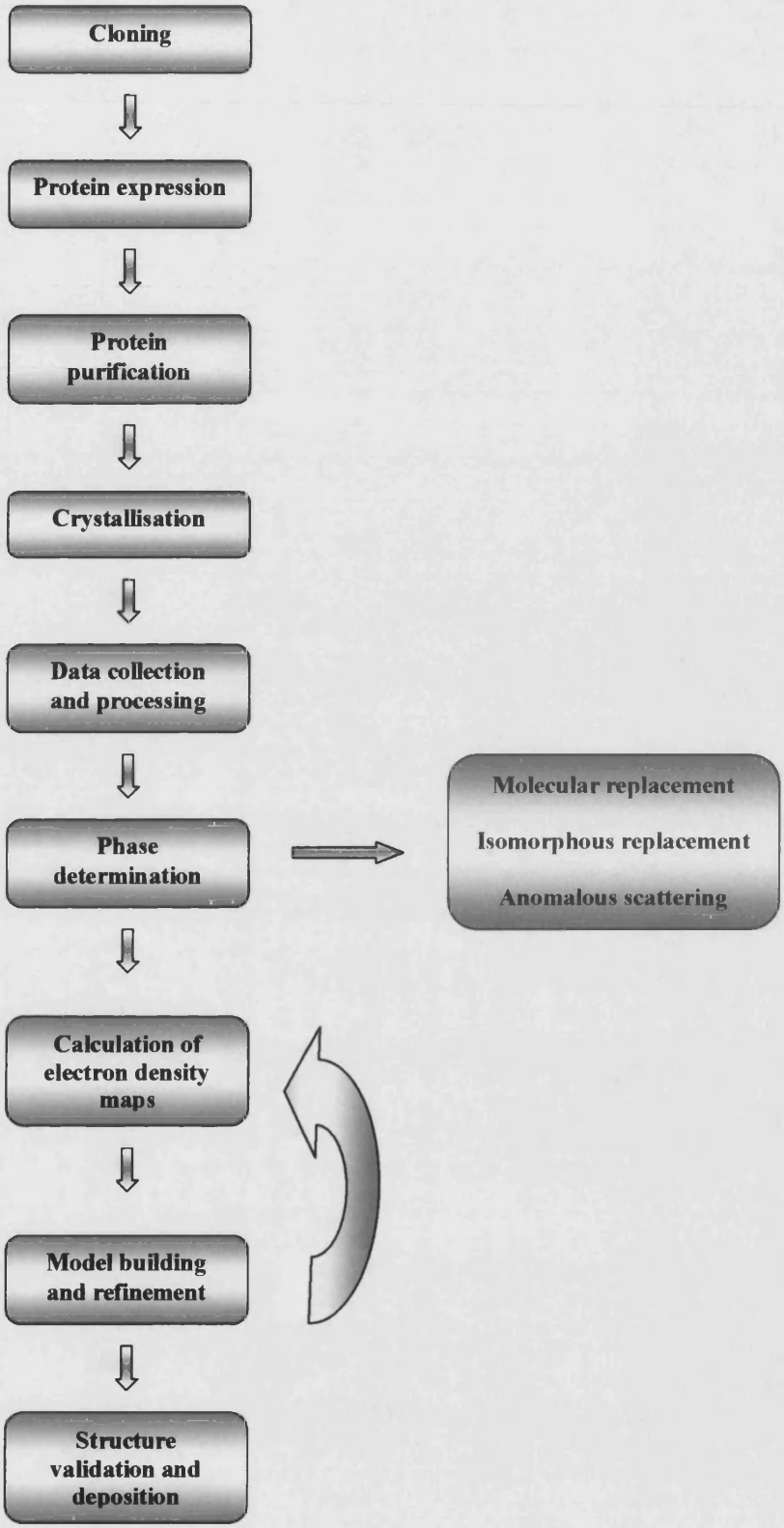


Figure 1.1: Flow diagram of the steps involved in the determination of macromolecular structures.

1.2 Cloning

This is the first step for the production of protein which will be subsequently used in crystallographic experiments. Identification of the gene of interest followed by its amplification is a process known as cloning. Initially, the DNA sequence of interest is amplified through a process known as polymerase chain reaction (PCR). The amplified DNA sequence is then introduced into a suitable cloning vector, a 'DNA vehicle'. The choice of vector can affect following steps of protein expression and purification and should be thus done very carefully. Plasmids and bacteriophages are most commonly used for cloning, while bacterial artificial chromosomes (BACs; Shizuya *et al.*, 1992) and yeast artificial chromosomes (YACs; Foote *et al.*, 1992) are also available for this purpose. Vectors containing ready to use affinity tags are especially useful as they can greatly enhance subsequent steps. The presence of a specific promoter in the vector is essential for control of expression and is usually included along with multiple cloning sites and a selectable marker (e.g. antibiotic resistance) (Figure 1.2). Once the vector to be used is decided, both vector and DNA of interest are cleaved with specific enzymes, known as restriction enzymes that produce different types of 'ends'; enzymes that produce blunt or sticky ends are most commonly used. The vector and insert (gene of interest) are then ligated such that the vector now carries the gene of interest.

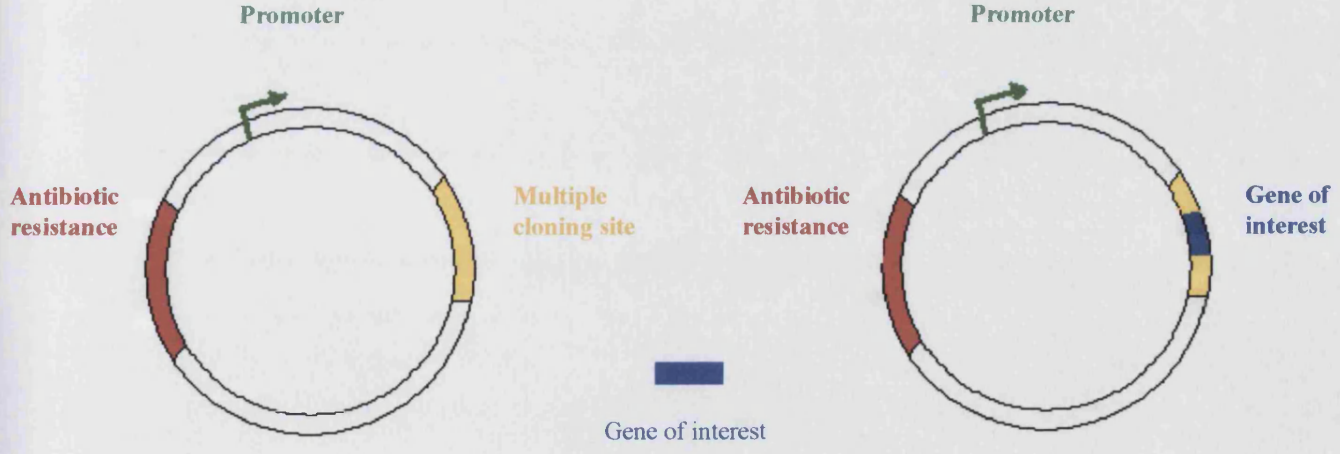


Figure 1.2: Gene cloning in the vector of choice. Shown here are the promoter, a selectable marker for antibiotic resistance and a multiple cloning site, where the gene of interest is inserted.

1.3 Protein expression

Protein expression *in vivo* (directly from wild type sources) can be expensive and impractical. For this purpose expression systems have been developed of both prokaryotic and eukaryotic origins. Prokaryotic recombinant protein expression systems provide ease of culture and rapid cell growth as well as controllable induction of protein expression using IPTG (isopropyl-beta-D-thiogalactopyranoside). One of the most commonly used prokaryotic expression systems is *E.coli*. More complex molecules that need post-translational modification in order to be fully active and acquire their native conformation are expressed using eukaryotic expression systems. These include yeast cells (e.g. *Pichia pastoris*), insect cells (e.g. *Drosophila melanogaster*) and mammalian cells (e.g. Baby hamster kidney cells). Conditions used in expression of the recombinant protein, such as temperature, pH and nutrients are of key importance and often need to be optimised in order to achieve better yield; as a lot of material is required for crystallisations, an expression method that yields significant amounts of protein is required.

1.4 Protein purification

Protein purification is one of the most complex steps in obtaining homogeneous protein. Several types of chromatographic methods have been developed for protein purification. Most commonly more than one of these methods is required for obtaining pure protein. Protein purification methods include affinity chromatography, ion-exchange chromatography and size exclusion chromatography. Affinity chromatography is one of the most widely used methods for protein purification, and is especially useful for proteins that have been expressed with a fusion tag. Most commonly used tags include hexa-histidine (His_6), glutathione S-transferase (GST) and maltose binding protein (MBP). The principle of affinity chromatography is the affinity of a biospecific ligand immobilised covalently on the chromatographic bed material for which the sample of interest has affinity. Sample and ligand will bind due to affinity interactions, while any unbound material will be washed off. It is essential that methods for desorbing the bound sample of interest are available. Ion-exchange chromatography separates proteins according to the charge of the protein. This type of chromatography makes use of the

reversible adsorption of charged solute molecules to an immobilised ion matrix of opposite charge. Two types of exchangers are available: anion exchangers, which are negatively charged and cation exchangers, which are positively charged. Bound protein is eluted using a salt gradient. Size exclusion chromatography separates proteins according to their molecular weight. A porous gel matrix is used through which molecules in the sample loaded are separated according to differences in their size, with the larger molecules being eluted first from the column. Size exclusion chromatography is usually one of the last steps in protein purification and is not affected greatly by buffer composition as molecules do not bind to the gel matrix. Usually each purification step involves some loss of protein. This can, however, be minimised by selecting the most suitable methods and optimising each of the steps involved in obtaining pure protein. To ensure the purity of the protein of interest SDS-PAGE is carried out after each step of purification. Once sufficiently purified, the protein of interest is concentrated using filters with an appropriate molecular weight cutoff. Concentrated protein samples are tested again using SDS-PAGE for any remaining impurities or heterogeneity which can greatly affect subsequent steps.

1.5 Protein crystallisation

Crystals are regular arrays of molecules with a repeating pattern that extends in all three spatial dimensions (see Figure 1.5). Obtaining crystals of good diffracting quality is one of the most difficult stages in structure determination. This is mainly due to the number of variables that are involved in crystallisation a few of them being protein concentration and homogeneity, pH, temperature, salt and precipitant concentration. There are two steps in protein crystallisation: *nucleation* and *crystal growth*. Nucleation refers to the formation of the first aggregates of molecules that act as nuclei for further crystal growth. Once the first nuclei are formed crystal growth can result from supersaturation in the medium where the first aggregates appeared. This process of controlled precipitation of the protein is illustrated in Figure 1.3.

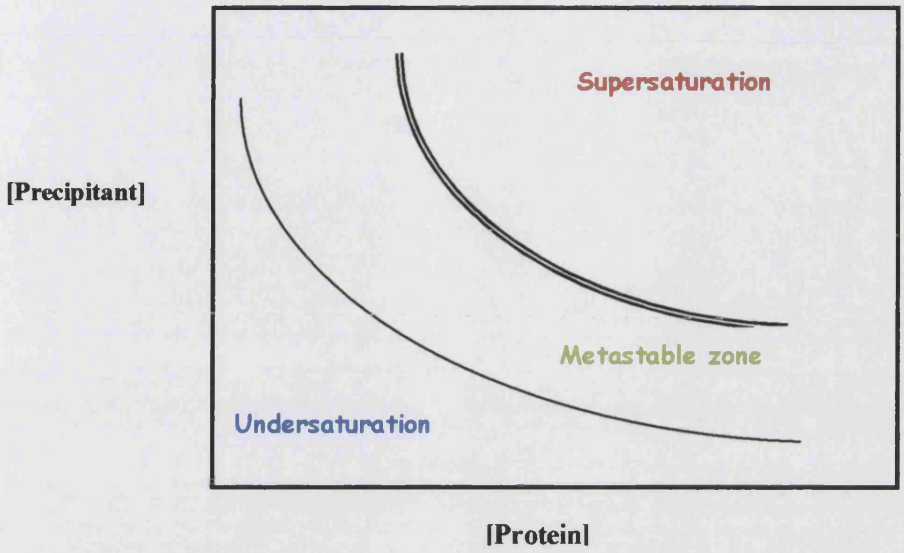


Figure 1.3: Phase diagram for protein crystallisation. The x-axis represents the protein concentration, while the y-axis represents the precipitant concentration. In low protein and precipitant concentrations, the solution is undersaturated, while the use of high precipitant and protein concentrations usually lead to the opposite extreme. Reaching the metastable zone is essential for nucleation which can lead to crystal growth.

Different methods can be employed for protein crystallisation. These include the batch method, dialysis button method and vapour diffusion method. Hanging drop vapour diffusion is the most commonly used method for protein crystallisation; Figure 1.4. This method is advantageous in that it requires small amounts of protein sample. In the hanging drop method, a drop of about 1-3µl of protein mixed with an equal amount of solution (mother liquor) that contains all the precipitants is placed on a siliconised coverslip. The coverslip is then placed on top of a reservoir containing mother liquor and is sealed with grease. The two solutions tend towards equilibrium and this causes vapour diffusion from the protein solution to the reservoir as the latter is of higher concentration. Vapour diffusion results in a more concentrated protein solution through which supersaturation of the protein solution is achieved. This in turn favours nucleation and crystal growth.

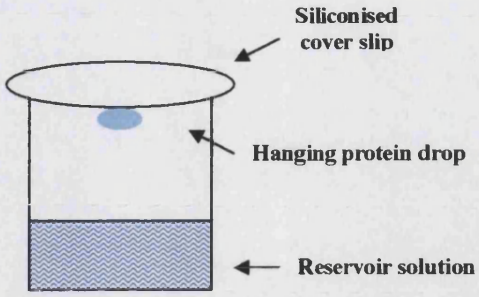


Figure 1.4: Hanging drop vapour diffusion method for protein crystallisation. The reservoir contains the mother liquor and sealed on top of the reservoir is the drop containing the protein sample mixed with reservoir solution. The higher concentration of precipitant in the reservoir solution causes vapour diffusion from the drop to the reservoir, resulting in supersaturation of the protein sample.

As there is a wide range of precipitants, salts and buffers to be screened initial crystallisation trials include the use of pre-made screens which cover a range of these conditions. Initial hits usually require further optimisation in order to obtain good quality diffracting crystals.

1.6 Crystals and symmetry

Crystals are highly ordered three dimensional arrays of molecules with a pattern that is repeated at regular intervals. This repeating motif is the fundamental block of a crystal and is known as the **unit cell**. A unit cell is defined by its three edges a, b, c and three angles α, the angle between b and c, β, the angle between a and c and γ, the angle between a and b; Figure 1.5. The unit cell contains all the information required to elucidate the structure of the macromolecule crystallised. The asymmetric unit is the smallest repeating object in a crystal lattice. Symmetry operations on contents of the asymmetric unit generate a unit cell and subsequently the rest of the crystal; Figure 1.5. Each asymmetric unit may consist of one molecule, one subunit or a complex of more than one molecule.

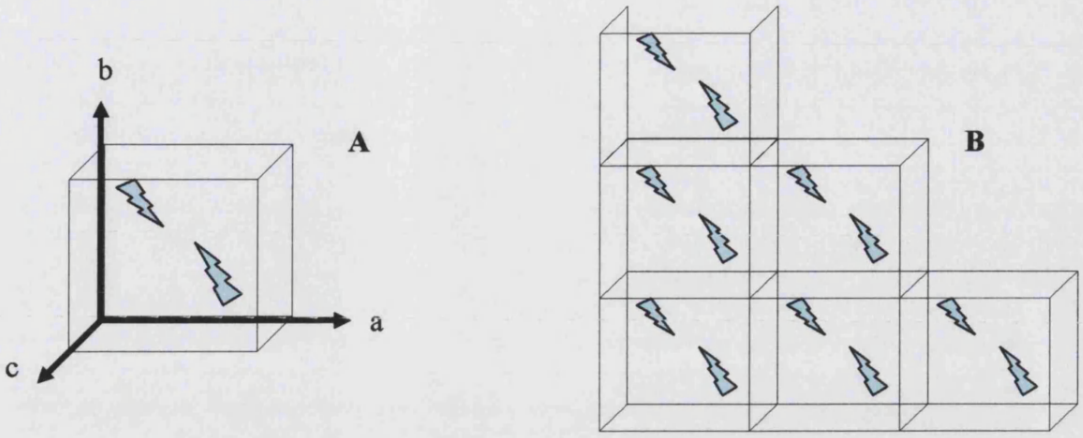


Figure 1.5: Schematic representation of the unit cell (**A**), with three edges a, b and c and three angles α, β and γ. Application of symmetry operations on the unit cell generates the crystal lattice (**B**).

Based on their symmetry, crystals can be grouped into seven crystal systems, starting from triclinic, which is the lowest order of symmetry (Table 1.1). These seven crystal systems are also known as the *primitive lattices*. Bravais described another seven *non-primitive lattices*. The seven primitive lattices along with the seven non-primitive lattices are collectively known as the Bravais lattices (Blundell and Johnson, 1976).

Symmetry operations like rotation, mirror plane and inversion that pass through a single point form the point group symmetry. These symmetry operators about a point constitute the 32 point groups. Further symmetry can be observed in space groups, which are the two-dimensional representations of a three-dimensional object (Blundell and Johnson, 1976). Symmetry operators in space groups include translations, screw axes and glide planes. There are 230 space groups, which arise from the 32 point groups and the fourteen Bravais lattices. Protein macromolecules however consisting of amino acids, can only crystallise in one of the 65 'allowed' space groups as mirror planes of amino acids do not exist in nature. Amino acids (except for glycine) have a chiral carbon atom adjacent to the carboxyl group (CO_2^-) and those found in proteins occur in the L-configuration about the chiral carbon atom. D-amino acids are not naturally found in proteins and are not involved in the metabolic pathways of eukaryotic organisms.

Table 1.1: Spacegroups, Bravais lattice types and conditions required for categorisation.

Name	*Bravais Lattice types*	*Conditions*
Triclinic	P	$a \neq b \neq c$; $\alpha \neq \beta \neq \gamma$
Monoclinic	P, C	$a \neq b \neq c$; $\alpha=\gamma=90° \neq \beta$
Orthorhombic	P, C, I, F	$a \neq b \neq c$; $\alpha=\beta=\gamma=90°$
Tetragonal	P, I	$a=b \neq c$; $\alpha=\beta=\gamma=90°$
Trigonal	P	$a=b \neq c$; $\alpha=\beta=90°$, $\gamma=120°$
	or R	$a=b=c$; $\alpha=\beta=\gamma<120°. \neq 90°$
Hexagonal	P	$a=b \neq c$; $\alpha=\beta=90°$, $\gamma=120°$
Cubic	P, I, F	$a=b=c$; $\alpha=\beta=\gamma=90°$

1.7 Bragg's law

W. L. Bragg (Bragg, 1913) visualised the scattering of X-rays by a crystal and identified its resemblance to ordinary reflection. According to Bragg's law, constructive interference between rays scattered from successive planes in the crystal will only take place if the path difference between the rays is equivalent to an integral number of wavelengths; Figure 1.6.

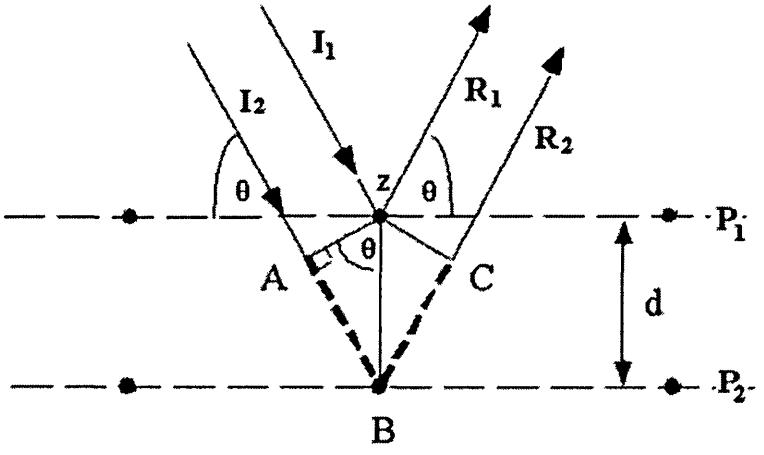


Figure 1.6: Bragg's law. Where P_1 and P_2 are two successive crystal planes, separated by distance d. Where I_1, I_2, R_1 and R_2 are the incoming and reflected X-rays respectively, at an angle θ of the crystal planes.

Where, I_1, R_1 and I_2, R_2 is the incident and the reflected beam by planes P_1 and P_2 respectively. Where θ is the angle of the incident and the reflected beam. The distance between the two planes of the crystal lattice is d. The law of diffraction states that two waves are in phase when their path difference is an integral multiple of the wavelength, λ (units of wavelength λ is Å=10^{-8}cm). As seen in Figure 1.6, the path difference between the two waves is equal to AC+CB. In order for the two waves to be in phase AC+CB has to be an integral multiple of the wavelength. Hence,

$$AB+BC=2AC=n\lambda \quad [1]$$

Trigonometrically however, AC can also be expressed as:

$$AB=d\sin\theta \quad [2]$$

Combining equations *[1]* and *[2]*:

$$n\lambda=2d\sin\theta \quad \text{(Bragg's law)}$$

Bragg's law allows the prediction of the position of any diffracted ray in space. Considering that the wavelength of the X-rays going in to the crystal is known and the angle θ of the diffracted X-rays coming out of the crystal can be measured, the inter-planar spacing can be calculated.

1.8 Ewald sphere and reciprocal space

According to Bragg's law the closer the planes in a crystal lattice the wider the diffraction observed in the diffraction pattern, while lattices with larger interplanar distance produce a closer diffraction pattern. Hence the diffraction observed is in reciprocal space rather in real space (instead of interplanar distance d in real space, it is 1/d in the reciprocal space). Not all points in a crystal lattice will contribute in the production of a diffraction pattern. A geometric construction by Ewald (Ewald, 1921) aids in the visualisation of the reciprocal space and the diffraction geometry; Figure 1.7.

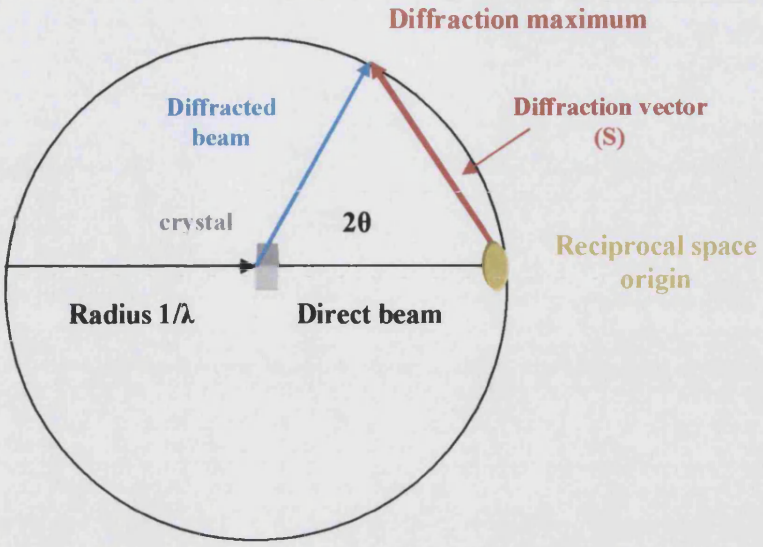


Figure 1.7: Ewald sphere construction is a three-dimensional representation of Bragg's law. It is a sphere of 1/λ radius, where the crystal is located at the centre and rotated perpendicular to the incoming X-ray beam. Diffraction is produced when the angle of the diffracted beam is equal to 2θ and Bragg's law is satisfied.

Ewald sphere has a radius of 1/λ. The crystal is located at the center of the sphere. The incoming beam terminates at the origin of the reciprocal space. A diffraction "spot" will only be produced by those points that lie in the circumference of the sphere, hence reciprocal lattice points. When the angle between the incoming and outgoing beam will be equal to two times theta and the length of the diffraction vector S will be equal to 1/d, Bragg's law will be satisfied. These points will thus contribute in the production of a diffraction pattern. Each point observed in the diffraction pattern is assigned a set of three indices which are used to define the orientation of crystallographic planes within a crystal. These are known as the *Miller* indices *h*, *k*, *l*. The origin of the reciprocal space is taken as the central beam position as seen in a diffraction pattern. *Miller* indices for the origin of the reciprocal space are *hkl* = 0, 0, 0 or *h*=0, *k*=0 and *l*=0.

1.9 Diffraction pattern and resolution

When a crystal is "hit" by X-rays and Bragg's law is satisfied a diffraction pattern will be produced; Figure 1.8. This pattern can be used for deriving preliminary information required for structure elucidation. A single diffraction pattern allows the determination of the unit cell, the spacegroup and the resolution. Parameters such as mosaicity of the crystal, the exposure time required for a complete data set and the crystal-to-detector distance can be decided upon observation of a diffraction pattern. The pattern produced by a diffracting crystal provides information about the position of atoms given by the *Miller* indices and the intensity of each "spot" at that particular position, I_{hkl}.

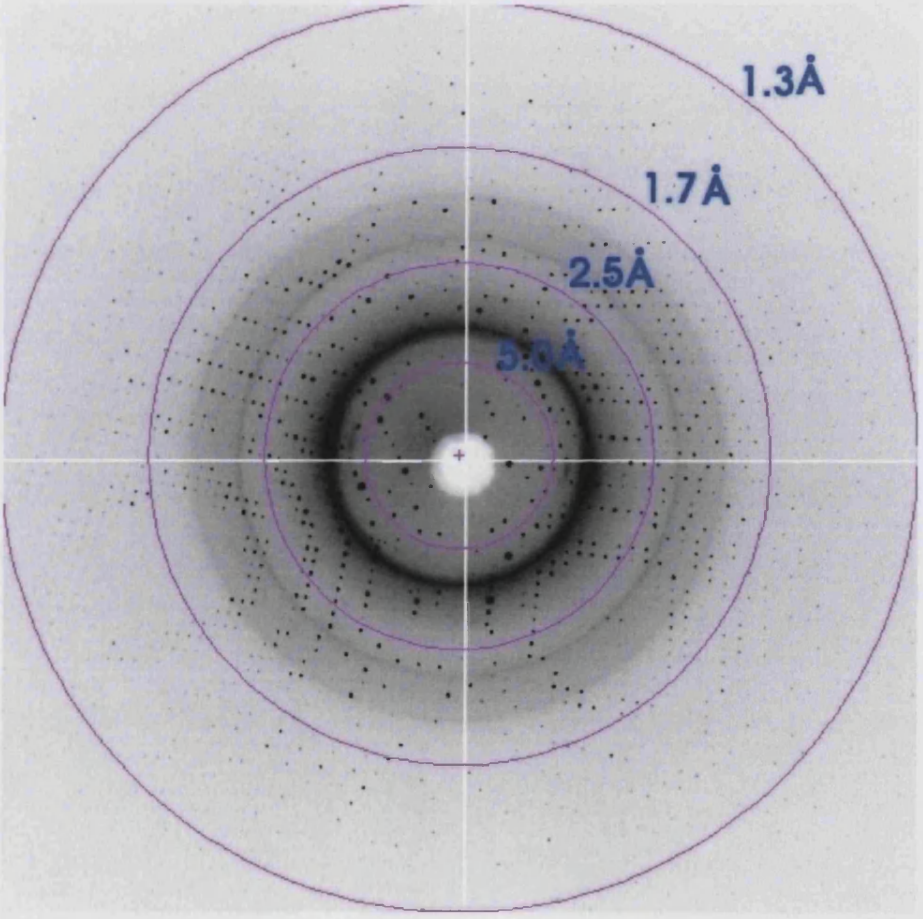


Figure 1.8: Diffraction pattern produced by eosinophil derived neurotoxin crystal (EDN; unpublished data). Shown in pink are the resolution shells; the furthest from the centre of the beam, the higher the resolution.

A very important parameter in data collection is resolution as the level of detail that will be obtained for the structure of a macromolecule depends on the resolution of the collected data (Table 1.2; Blundell and Johnson, 1976). Resolution is used to describe the distance between two very proximal spots observed on the diffraction image. Resolution is measured in Angstroms, which is a unit of length, equal to 10^{-8}cm.

Table 1.2: Effect of resolution on the structural information obtained by a crystal.

Resolution (Å)	Structural information obtained
6.0	Outline of the molecule, secondary structural features such as helices and strands can be identified.
3.0	Course of the polypeptide chain can be traced and topology of the folding can be established. With the aid of amino acid sequence, it is possible to place the side chains within the electron density map.
2.0	Main chain conformations can be established with great accuracy. Details of side chains conformations, bound water molecules, metal ions and cofactors can be identified.
1.5	Individual atoms are almost resolved. It is possible to make out the solvent structure.
1.0	Hydrogen atoms may become visible.

The technology used for X-ray data collection and processing has improved immensely. Laboratory sources such as sealed tube and rotating anode devices accelerate electrons towards a metal source, usually copper, that then emit X-rays at a characteristic wavelength: CuKα, 1.54Å wavelength; if shorter wavelengths are required molybdenum can be used as an anode (MoKα, 0.71Å). The X-rays produced are filtered through a crystal monochromator and then focused by curved mirrors. Laboratory sources are about 100-1000 times less powerful than synchrotron radiation sources; laboratory sources accelerate electrons to around 45-50kV while in synchrotrons accelerated electrons reach

a final energy of 1GeV to 8GeV. The development of synchrotrons has greatly enhanced the process of data collection; instead of the low intensity beam (proportional to the number of electrons emitted at any given time) emitted by the in-house X-ray sources, synchrotrons produce electromagnetic radiation. Detector improvement has also contributed immensely to the better quality and fast data collection. Nowadays, use of CCD (Charged Couple Device) provides better count rate, better spatial resolution and lower noise detection.

1.10 Data collection

Data collection is a critical step in the process of structure elucidation. Several limiting factors can affect data collection, the most important being the quality of the crystal and hence the quality of the collected data. Parameters that need to be considered for data collection include *exposure time*, *crystal mosaicity*, *crystal-to-detector distance* and *oscillation*. Longer *exposure time* is required for crystals that produce weak high resolution reflections. Increasing the exposure time can increase the intensity of the spots collected. This however can result to saturation of some reflections, seen as spots in the diffraction pattern, and can also increase the radiation damage caused to the crystal. In such a case data of lower resolution can be collected, where the exposure time is reduced. *Crystal mosaicity* refers to the internal disorder of a crystal; this occurs as not all molecules that form the crystal lattice share the same orientation. Mosaicity causes the X-ray beam to become spread into a narrow cone and can be one of the reasons for some reflections to be recorded as a spread over two or more images; these are termed partial reflections. Reflections that are recorded in a single image are termed full reflections. Mosaicity can be easily detected from the diffraction pattern as spots look broadened. High mosaicity can result in overlapping spots and data loss. *Crystal to detector distance* determines the resolution of the collected data. It also affects spot size in the diffraction pattern. The further the detector from the crystal the lower the resolution and the background noise from the scattered radiation from the crystal (Blow, 2002). The angle of *oscillation* required in order to collect a complete data set depends on the symmetry of the unit cell of the crystal. The higher the symmetry of the crystal, the less angle of rotation is required in order to collect a complete data set. Usually more reflections are

collected in order to decrease the signal-to-noise ratio and increase the redundancy of the data collected.

Data collection can take place either at room temperature or under cryogenic conditions. Room temperature data collection involves the use of thin walled glass or quartz capillaries where the crystal is drawn by suction; mother liquor is also drawn in the capillary but excess liquid is removed. The tube is sealed off with wax so as to prevent the crystal from drying out. Data collection under cryogenic conditions has proven to produce higher quality data compared to those collected at room temperature. In this case of data collection the macromolecular crystal is usually dipped in a solution containing mother liquor and some *cryoprotecting* agent; *cryoprotecting agents* most commonly include polyols, such as glycerol and polyethylene glycols of various molecular weights. The crystal is then mounted on to a loop where it is held by means of surface tension and it is flash frozen in a stream of liquid nitrogen which solidifies the solution on the loop and holds the crystal stable and under low temperature. Cryogenic conditions generally protect the crystal from radiation damage; damage caused to the crystal by the absorption of the energy released by the X-rays, which increases the internal disorder of the crystal, leads to the production of free radicals and reduces the intensity of scattering (Blow, 2002). Radiation damage can lead to lower quality data as well as data and crystal loss; the use of cryogenic conditions however usually allows the collection of a full dataset from a single crystal.

1.11 Data processing – reduction and scaling

Data processing refers to the indexing of each reflection observed in the diffraction pattern and subsequently of all the diffraction patterns in the collected dataset. What is observed in the diffraction pattern is the intensity of the reflections measured; during data processing these intensities are converted to amplitudes. Data processing comprises of two steps: data reduction and scaling. Several automated computer programmes have been developed for data processing. These include MOSFLM (CCP4, 1994) and the HKL package (Otwinowski, 1993a; Otwinowski, 1993b). HKL package includes DENZO for

autoindexing, refinement and integration, XdisplayF for visualising each image and SCALEPACK for scaling the indexed dataset.

Data processing starts with autoindexing of the diffraction pattern, where the symmetry of the unit cell and the symmetry of the crystal are used to produce a list of lattice types on the basis of the symmetry rules in the International Tables for crystallography (Hahn, 1987) and a percentage of distortion. The space group that will be used as an input in the following steps of data processing is chosen as the best combination of highest symmetry and lowest percentage distortion. Data reduction also refines the crystal and detector parameters.

Scaling averages the processed data while accounting for errors that occur during data collection (e.g. crystal decay). Scaling output produces a list of statistics that help determine the quality of data collected, such as *percent completeness*, R_{sym} or R_{merge} and $I/\sigma I$. *Percent completeness* refers to the completeness of the data; the higher the percentage, the more information will be derived by data processing. R_{sym} gives the estimate between the measured and the observed intensities. When more than one dataset are scaled together, then this value is known as R_{merge}. $I/\sigma I$ is the error in measuring the intensity of a reflection. A value of I/σI above 2.0 is indicative of an acceptable resolution limit for data scaling.

1.12 Phase problem

Structure determination by X-ray crystallography aims at using diffraction data and translating it into distribution of the electron density in the unit cell. Calculation of electron density requires both amplitudes and phases of the recorded reflections. During diffraction data collection phases are lost and it is not possible to identify the atomic positions based only on the information obtained from the amplitudes. Hence, calculation of the phases from the diffraction images is not possible. This is known as the *phase problem.*

Structure factor can be defined as the effective number of scattering electrons and depends on the distribution of atoms in the unit cell and the scattering direction. The structure factor is given by the formula:

$$F_{hkl}=\Sigma f_j e^{2\pi i(hx+ky+zl)}$$

Where F_{hkl} is structure factor of the atom at position hkl

Where f_j is the atomic scattering factor (for any atom j)

From the above formula it can be seen that structure factor is the sum of scattered waves describing one particular reflection. The number of electrons per unit volume (V) is known as electron density $\rho(x,y,z)$ and is given by the formula:

$$\rho(x,y,z)=(1/V)\ \Sigma_{hkl} F_{hkl}\ e^{-2\pi i(hx+ky+lz)}$$

Where $\rho(x,y,z)$ is the electron density over volume V

Phases however cannot be measured directly; hence one of the two requirements for the calculation of the electron density is missing. Methods used for phase determination are discussed later in this chapter.

1.13 Methods for phase determination

There are three most commonly used methods to derive the phases (φ); molecular replacement, isomorphous replacement and anomalous scattering.

1.13.1 Molecular replacement

Molecular replacement (MR) method is one of the most popular methods for determination of a macromolecular structure when a structure which shares >30% sequence identity with the unknown protein target is available. Sequence homology and

quality of the search model are two very important parameters for a successful MR solution.

Two methods used for molecular replacement are the *maximum likelihood*-based method and the *Patterson*-based method. In *maximum likelihood* method a set of observations is given and the method detects the model parameters that are most consistent with the observations. It is a very accurate and sensitive method for the calculation of rotation and translation functions. Maximum likelihood method also accounts for different sources of errors in the model, such as coordinate errors, missing atoms and errors in the B-factors (thermal vibration factor). Molecular replacement programmes that make use of this method include Beast (Read, 2001) and Phaser (Read, 2001; Storoni *et al*, 2004).

Patterson based method requires a total of six parameters to be specified; three parameters (angles) are required for the rotational search and another three for the translational search (three-dimensional space). This method of molecular replacement breaks down the six parameters required by separately calculating the rotation and then the translation of the molecule. A Patterson function is a fourier transform of the measured intensities squared, but not phased. This produces a density map of the vectors between scattering objects in the unit cell, rather than an electron density map. Two types of vectors are used in Patterson function: inter-atomic and intra-atomic vectors. Intra-atomic vectors or self-vectors arise from the same molecule and are used to determine the relative *orientation* of individual atoms in the crystal by using the *rotation* function. Once the orientation is known, it can be used for the intermolecular vector calculation. Inter-atomic vectors or cross-vectors determine the relative *position* of these molecules and are used by the *translation* function (Rossmann and Blow, 1967). Programmes using the Patterson function based method for molecular replacement include MOLREP (Vagin and Teplyakov, 1997) and AMoRe (Navaza, 1994). The principle of molecular replacement is shown in Figure 1.9.

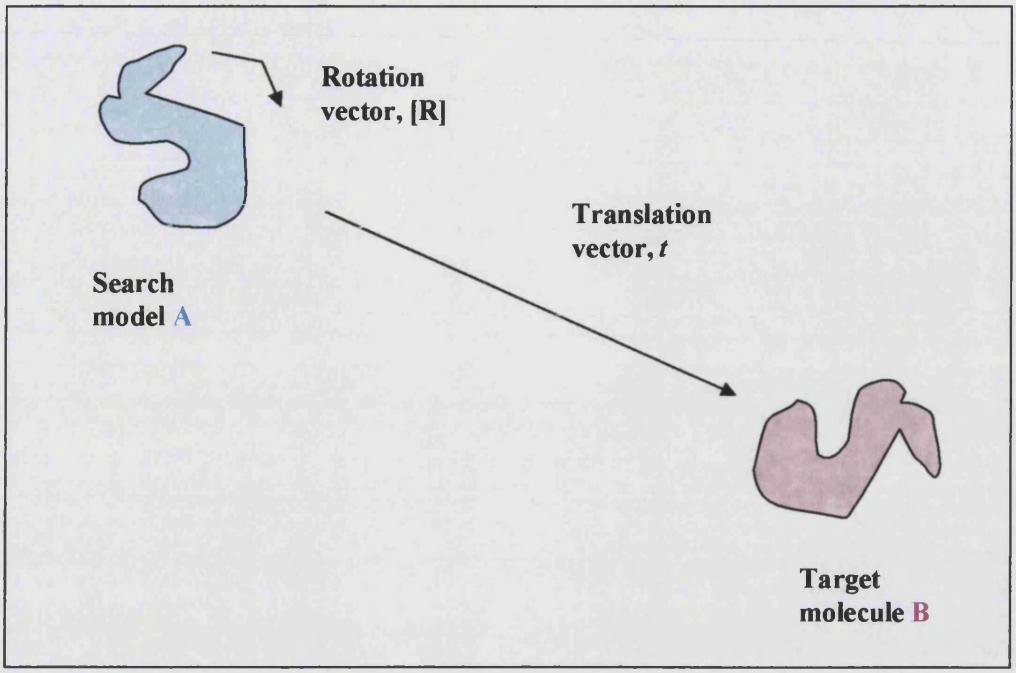


Figure 1.9: Principle of the molecular replacement method. The search model is illustrated in blue. Rotation R and translation of molecule A yields molecule B, which is the target molecule.

Molecular replacement programmes most often provide solutions for the rotation and translation of the molecule in the unit cell, each associated with a residual factor and a correlation coefficient. The residual factor is an estimate error between the observed and the calculated structure factors; the closer the agreement the lower the residual factor, the better the MR solution. The correlation coefficient describes the relationship of the two variables; observed and calculated structure factors. These parameters usually indicate whether a successful solution has been obtained.

1.13.2 Isomorphous replacement

Isomorphous replacement refers to the replacement of solvent molecules or light salt ions by the introduction of a heavy atom in the structure (Green *et al.*, 1954). This method exploits the principle that each atom in the molecule contributes independently to the

scattering factor. In Isomorphous replacement at least two crystals are required; the parent, which is that of the native protein and the derivative, which is the crystal containing the heavy atom(s). Heavy atoms can be introduced by soaking or by a chemical reaction. The introduced "heavy" atom of the derivative crystal will contribute disproportionately to the overall intensity. It inevitably disturbs the environment around the protein and usually forces some solvent molecules out of the crystal, which results in non-isomorphism. Data collection from both crystals (parent and derivative) and calculation of Patterson maps will lead to the identification of the location of the heavy atoms in the crystals and their contribution to the structure factor. Single Isomorphous Replacement (SIR) and Multiple Isomorphous Replacement (MIR) (Green *et al.*, 1954) are two methods used widely for the calculation of phases for a new protein. The use of one only derivative is usually not enough for the determination of phases. In practice two derivative crystals are usually required in order for the phase angle to be determined accurately.

1.13.3 Anomalous scattering

The method of anomalous scattering resembles the method of isomorphous replacement but differs in that heavy atoms, integral to the protein under study are used in order to exploit differences in scattered intensities. This method requires the use of tunable wavelength and is very popular for use at the synchrotrons. In anomalous scattering method, Friedel's law (Friedel, 1913), according to which $F_{hkl}=F_{-h-k-l}$ (where F_{hkl} and F_{-h-k-l} are Braggs reflections that have the same magnitude but phases of opposite signs) is breached. Tuning the wavelength close to the energy of electronic transition of the anomalous scatterer (absorption edge) can bring the latter to an excited state. In this case the radiation energy of the excited atom is not in phase with the incident beam, while the scattering of the "normal" component is reduced. As the wavelength is tuned and the crystal remains the same, this method has the advantage of accuracy. Anomalous scattering can take place either as Single-wavelength Anomalous Dispersion (SAD) or via Multiple-wavelength Anomalous Dispersion (Hendrickson, 1991; MAD). An even more powerful method for phase determination is the combination of Isomorphous replacement with anomalous scattering, if the anomalous scattering atoms are the same as

the heavy atoms used in an Isomorphous replacement experiment. Single wavelength Isomorphous replacement and anomalous scattering (SIRAS) and Multiple wavelength Isomorphous replacement and anomalous scattering (MIRAS) are two ways in which Isomorphous replacement and anomalous scattering can be combined to derive more information about the phases.

1.14 Model building and refinement

Once the phases have been obtained by molecular replacement, isomorphous replacement or anomalous scattering refining the agreement between the diffracted data and the calculated phases is the next step for structure elucidation. The structure image is in the form of an electron density map, electron density being the number of electrons per unit volume (electron density is the fourier transformation of structure factors). Electron density maps provide a wealth of information not only for the structure but for the progress of refinement as well. Difference density maps are calculated after every step of refinement. Two types of maps are most commonly used; F_o-F_c and $2F_o$-F_c maps. F_o-F_c electron density maps are difference maps where the observed and calculated structure factors have been given the same weight. Positive regions in the F_o-F_c maps will signify features that exist but have not been represented by the model. Negative regions in an F_o-F_c map will represent features that are present in the model but not supported by the observations. Double difference map, or $2F_o$-F_c, where the observed structure factors have double the contribution of the calculated structure factors will more strongly represent uninterpreted features in the structure. The initial electron density maps will inevitably have several areas that need correction. It is essential though to check that significant agreement exists between the model and the data (electron density map) before proceeding to the refinement of the structure. In cases where molecular replacement has been used for phase determination, the backbone chain of the search model used should agree and follow the mainchain of the electron density map, even at low resolutions. When a novel protein is under study and the phases have been determined by isomorphous replacement or anomalous scattering it is essential to take extra care in fitting the protein sequence (i.e. amino acids) into the electron density map.

The agreement between the observed and calculated structure factors can be significantly improved with iterative cycles of refinement and manual model building. Refinement programs that are most commonly used are REFMAC (Murshudov *et al.*, 1997) and CNS (Brünger *et al.*, 1998). Several parameters are refined during this process. These include positional parameters (coordinates x, y, z) and thermal parameters (atomic thermal factors). Satisfactory refinement of these parameters depends on the available number of observations. The observation: parameter ratio can in some cases prove to be a limiting factor in refinement. If such a ratio is low then not enough information is available from the collected data in order to account for all the parameters that need to be refined. The observation to parameter ratio is usually not high enough to account for the refinement of all parameters for each atom. Hence restraints and constraints are used in the refinement process. In restrained refinement parameters are allowed some freedom of movement. A type of restrained refinement is B-factor refinement, which refines the individual atomic thermal vibration. Restrained refinement improves the observation to parameter ratio, improving thus the refinement. In constrained refinement parameters are treated as rigid bodies and no distortion is allowed; no deviation from a user-specified value. A parameter where constraints can be applied is the occupancy, to which a certain value is given and no distortion from this value is allowed during refinement. Constrained refinement reduces the number of parameters requiring refinement, improving the observation to parameter ratio.

There are two methods for refinement: least squares refinement (LSQ) and maximum likelihood refinement (MLH). In least squares refinement the best values for a set of parameters are those which minimise the sums of the squares of the weighted differences between the observed and the calculated values. LSQ method for refinement is used by REFMAC (Murshudov *et al.*, 1997). The goal of the maximum likelihood method is to determine the probability of a set of measurements to occur, given the model, and estimates of its errors and of errors in the measured intensities. Errors that need to be accounted for in the maximum likelihood method include coordinate errors, missing atoms and errors in the B-factors. The maximum likelihood method is used in CNS (Brünger *et al.*, 1998) and TNT refinement (Tronrud, 1997).

Improvement of the quality of the electron density map is the result of improvement of the model via model building and refinement. One such refinement process is simulated annealing. Annealing is a physical process in which a solid is heated to a temperature high enough to make it liquid and then cooled down slowly so as for each and every atom to reach the lowest energy state. In simulated annealing the model is heated to a temperature defined by the user (usually 3,000-5,000K) and then it is allowed to cool down in small steps so that every atom can reach its minimum energy state. The quality of the model is thus improved in consecutive steps of refinement and the calculated atomic parameters being refined (spatial parameters and atomic thermal factor) are brought closer to the observed values.

Automated refinement corrects errors in the model and improves the agreement between observed and calculated structure factors. Gross errors however in the model cannot be corrected by refinement programs. Manual model building is essential in the improvement and interpretation of the electron density map. Through model building it is possible to check whether all electron density has been accounted for and whether the best possible interpretation has been given to the electron density map. Model building involves monitoring the refined model and electron density maps in molecular graphics programs such as COOT (Emsley and Cowtan, 2004) and O (Jones *et al.*, 1991). Examination of the difference density (positive and negative) will indicate whether refinement is proceeding in the correct direction. Electron density for solvent atoms, ions and any ligands present should also be accounted for, nearing the end of the refinement in order to avoid misinterpretation or bias the electron density map. Programs for automated identification of solvent atoms are available (e.g. WATERPICK in CNS; Brünger *et al.*, 1998) while ions and ligands have to be defined by the user. A database where several popular ligands are deposited is the HIC-Up web-server (Kleywegt and Jones, 1998).

The measure of a refining structure is given by the reliability index or R-factor. As refinement progresses the agreement between the structure factors Fs, calculated F_{calc} and observed F_{obs}, improves and as a result the R-factor drops. The R-factor is calculated from the following formula:

$$R_{cryst} = \frac{\Sigma_{hkl} ||F_{obs}| - k|F_{calc}||}{\Sigma_{hkl} |F_{obs}|}$$

Where, Σ_{hkl} is the sum of the differences of $||F_o| - |F_c||$ over h, k, l

Where, $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes

Where k is the scale factor

To avoid overfitting the data during crystallographic refinement, Alex Brünger (1992) introduced another parameter, R_{free}, which states the quality of a structure and is a cross-validation parameter. Prior to refinement the total number of reflections obtained by data collection is separated into two groups prior refinement; a working set and a test set. Working set indicates the agreement, R_{cryst}, between the observed and calculated data. It decreases when favourable changes are made to the model. The test set, R_{free}, consists of around 750-1000 reflections (~5% of the total number of reflections) picked randomly and separated from the total number of unique reflections. R_{free} is calculated against the data instead of against the model; hence there is no model-bias in the refinement of the test set. If the errors that occur during model refinement are only statistical and the model is correct then both R-factors should decrease. Iterative cycles of model building and refinement come to an end when all the electron density has been interpreted or there is no observable change in the R-factors.

1.15 Structure validation, analysis and deposition

Validation is a necessary step for the assessment of the refined structure. Several programs are available for structure validation. The most commonly used program to asses the geometry of the refined structure is PROCHECK (Laskowski *et al.*, 1993). PROCHECK provides a Ramachandran plot (Figure 1.10; Ramachandran *et al.*, 1963) which indicates the number of residues in the allowed and disallowed conformations based on φ and ψ angles defining the polypeptide backbone. Ideally, a refined structure should have 0% of the residues in disallowed conformations with the vast majority in the

most favourable region. This program also produces a list of distorted geometry in terms of bond angles, bond lengths and planar groups. Outliers noticed in the Ramachandran plot or in the distorted geometry output files have to be checked carefully. Divergence from the expected values can either be because of inaccuracy or they might indicate a genuine feature of the structure (e.g. cis-peptide bonds, crystal packing, ligand contact).

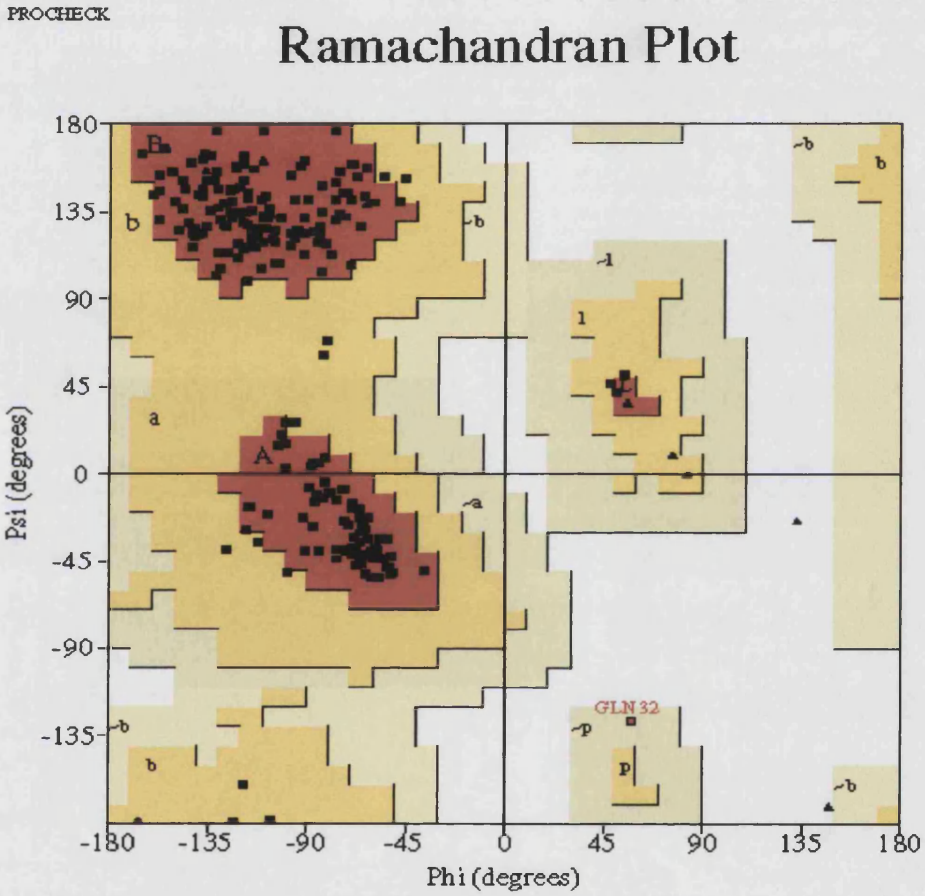


Figure 1.10: Representation of a Ramachandran plot. The x-axis represents the φ angles, between C-N while the y-axis represents the ψ angles, between C-C. The red regions of the plot represent the secondary structure elements with top left corner representing the beta sheet, the bottom left corner the right-handed alpha helix and the top right corner the left handed alpha helix. Regions shown in yellow represent the additionally allowed regions. Regions shown in beige are the generously allowed regions while white regions represent the disallowed amino acid conformations. ▲ represent glycines while ■ represent all other amino acids.

Similarly, model_stats in CNS (Brünger *et al.*, 1998) provides an output list file with information about the geometry and violations, close contacts and crystal contacts. Root mean square deviation (rmsd) values given for bond lengths and bond angles are parameters indicative for the quality of the geometry of the refined structure. Close contacts and crystal contacts have to be examined in order to correct any errors in the refined structure.

MOLPROBITY (Davis *et al.*, 2004) is a web-based program, which checks for close contacts and interactions amongst residues in the refined structure. A crystal structure of good quality would ideally have no close contacts. If any close contacts are detected in the refined structure the program produces a list of alternative amino acid side chain conformations that can be checked in order to improve the quality of the structure.

Only after validation can the interpretation of the structure, in the form of structure analysis, begin. Several programs are available for examining different aspects of the structure. For assigning the secondary structure elements in the structure DSSP (Kabsch and Sander, 1983) is the program most commonly used. CONTACT from the CCP4 suite (CCP4, 1994) produces a list of all interactions (hydrogen bonds and van der Waal interactions) detected in the structure. This is particularly useful if ions and ligands are present in the structure as well as for examining the interactions of certain residues with others, or with solvent atoms. HBPLUS (McDonald and Thornton, 1994) is the most commonly used program for the detection of hydrogen bonds as it calculates hydrogen bond lengths and donor to acceptor angles. Burried_surface (Lee and Richards, 1971) is a very useful program when complexes or ligands are examined in a structure. It provides the total surface area of a single chain molecule and upon formation of a complex (or when a ligand is bound to a specific site on the molecule) it provides the surface area that becomes buried. Several other programs can be used if more specific areas are to be analysed when interpreting the structure and possible function of a macromolecule.

A structure that has been validated and analysed can then be deposited with the Research Collaboratory for Structural Bioinformatics Protein Data Bank. The PDB (Bernstein *et*

al., 1977) contains structures determined by NMR, X-ray crystallography and some deposited theoretical models. Structure deposition includes the deposition of atomic coordinates, structure factors, experimental details as well as details about data collection, processing and refinement.

Chapter - II -

Introduction to angiogenesis

2.1 Introduction

Angiogenesis can be defined as the sprouting of new capillaries from preexisting microvasculature. The process of angiogenesis is similar to that of vasculogenesis, the latter describing however the *de novo* formation of the vascular system. Angiogenesis is a key phenomenon for many physiological and pathological processes. Physiological angiogenesis occurs during embryogenesis, while for a healthy adult it occurs during wound healing and the female ovarian/menstrual cycle. Uncontrolled or dysfunctional angiogenesis has been shown to be associated with ischemia of the heart, atherosclerosis, rheumatoid arthritis and tumour growth and metastasis (Folkman, 1971; Arnold *et al.*, 1987; Folkman, 1995; Norrby, 1997; Folkman, 1998). Control of angiogenesis occurs *via* a signal network of activators and repressors, which act as 'on' and 'off' switches. Activators of angiogenesis are known as angiogenesis-stimulating factors (angiogenic factors) while repressors of angiogenesis are known as angiogenesis inhibitors (angiostatins).

Endothelial cells are essential for angiogenesis. They are normally quiescent but become stimulated during an angiogenic response leading to degradation of the basement membrane and chemotactic migration. Ultimately, they undergo division and form functioning capillaries. These events are all essential for recovery of homeostasis.

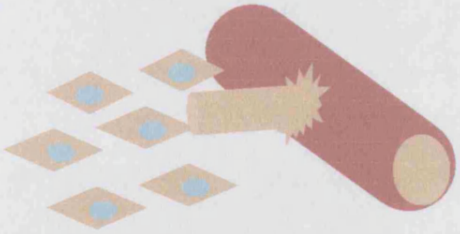
2.2 Physiological angiogenesis

Physiological angiogenesis occurs as a response to inflammatory cells or injury, where angiogenic growth factors such as vascular endothelial growth factor (VEGF), fibroblast growth factor (FGF), angiogenin and platelet-derived endothelial cell growth factor/ thymidine phosphorylase (PD-ECGF/TP), are released. The vessel wall of mature capillaries consists of a lining of endothelial cells, a basement membrane and a layer of pericytes, which surround the endothelium. The released angiogenic factors bind to receptors on the endothelial cell lining and stimulate them. This leads to endothelial cell growth and to the secretion of proteases, such as the matrix metalloproteases (MMPs), which digest the basement membrane surrounding the vessel and allow the endothelial

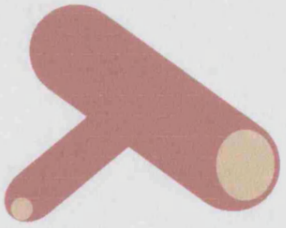
cell projections to migrate towards the diseased or injured tissue. The tissue surrounding the newly formed blood vessel is remodeled by MMPs, while pericytes form a structural support network, essential for stabilising the new capillary. Individual capillaries are linked together to form blood vessel loops through which blood circulates. Once the new vessel has been formed, blood flow initiates. The process of angiogenesis as a response to a stimulus stops when a balance between the positive and negative regulators of the process has been restored. A representation of the steps involved in angiogenesis is given in Figure 2.1 (Folkman, 2007).

A. Blood vessel, coloured in pink and the endothelial cell lining coloured in cream.

B. Upon the release of an angiogenic factor the endothelial cell lining begins to beak away and endothelial cells are released.



C. Endothelial cells migrate and sprout to form a new capillary (newly formed capillary shown in cream).



D. Endothelial cells attract pericytes, which form a structural support network, completing in this way the formation of a new blood vessel.

Figure 2.1: Schematic representation of the steps (**A-D**) involved in an angiogenic response.

2.3 Pathological angiogenesis

Dysfunctional and uncontrolled angiogenesis has been shown to be involved in many pathological conditions. Increase in the expression of angiogenesis activators triggers leads to excessive angiogenesis, responsible for diseases such as cancer, psoriasis and rheumatoid arthritis (Figure 2.2). In such cases newly formed blood vessels provide nutrients and oxygen to diseased tissues allowing and promoting the diseased state. Insufficient angiogenesis occurs in diseases such as coronary artery disease, stroke and delayed wound healing (Figure 2.2) and is caused by an increase in the expression of angiogenesis inhibitors. This in turn leads to inadequate blood vessel growth and hindrance of blood circulation.

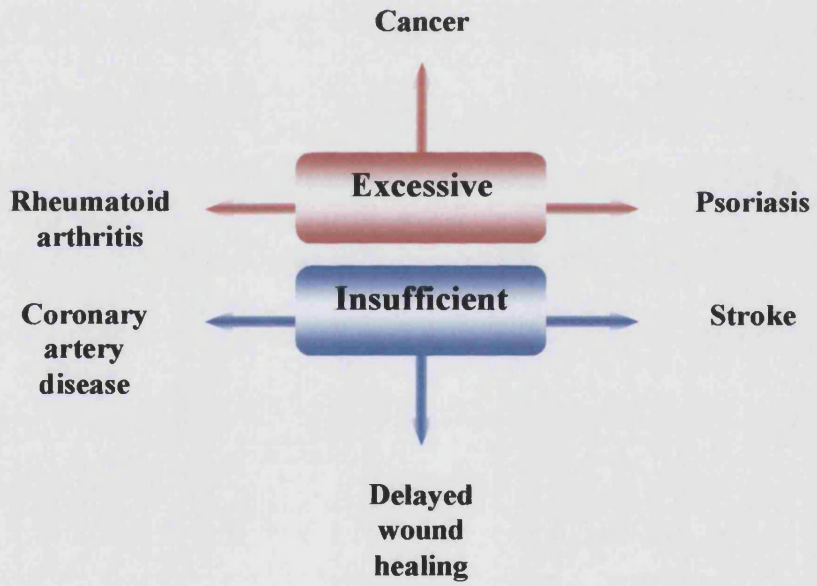


Figure 2.2: Pathological angiogenesis, excessive or insufficient, and its implication in several pathological conditions.

2.3.1 Tumour growth and metastasis

Tumour growth and metastasis has been shown to be dependent on angiogenesis in 1972 (Gimbrone *et al.*, 1972). In humans, most tumours in the early stages exist in an avascular, quiescent status, where they are restricted in their growth potential due to lack

of blood supply (Folkman, 2002). Such types of tumours can persist for months or years before becoming detectable. For tumours to increase in size and become detectable an angiogenic switch must be triggered. Once triggered the expression of angiogenesis activators increases and the balance between activators and inhibitors of is disrupted. In some cases tumours have been shown to express proangiogenic factors, such as vascular endothelial growth factor (VEGF; Kleespies *et al.,* 2005), that allow the angiogenic switch to take place. Formation of new blood vessels allows the transport of nutrients and oxygen to proliferating cancer cells, allowing tumour growth. Blood circulation provides an 'escape' route for cancer cells so that they can circulate and lodge on normal host tissue and other organs causing tumour metastases. Interestingly, apoptosis (programmed cell death), a process that occurs in normal tissue and maintains homeostasis, can be blocked in tumour cells. Mutations in specific genes and over-expression of proteins that suppress cell death allow tumour cell survival and promote tumour growth.

2.4 Regulation of angiogenesis

Regulation of angiogenesis is crucial and depends on the balance between activators and inhibitors. Angiogenesis stimulators include the VEGF, placenta growth factor (PlGF), FGF, angiogenin and PD-ECGF/TP (Milkiewicz *et al.,* 2006). PD-ECGF/TP has been shown to stimulate DNA synthesis and chemotaxis of endothelial cells *in vitro* and angiogenesis *in vivo* (Miyazono *et al.,* 1989). Its angiogenic activity has been shown to be dependent on the enzymatic activity, which is the reversible phosphorolysis of thymidine to thymine and 2-deoxy-D-ribose-1-phosphate. A key role is also attributed to proteins that are involved in the steps that lead to the formation of a new blood vessel. One class of such enzymes are the matrix metalloproteinases (MMPs), involved in the degradation of the extracellular matrix during the formation of a new blood vessel (Vu and Werb, 2000). Several types of MMPs with specific roles have been identified in humans, such as stromelysins, collagenases, gelatinases. The turnover of extracellular matrix in which these enzymes are involved is regulated by another class of enzymes, the tissue inhibitors of metalloproteinases TIMPs: the endogenous inhibitors of MMPs (Brew *et al.,* 2000; Wei *et al.,* 2003). In humans, four types of TIMPs have been identified and are responsible for the inhibition of the different types of MMPs. TIMPs have attracted a

lot of scientific interest as their role in inhibition of MMPs could be exploited for therapeutic purposes.

Balance between positive and negative regulators of angiogenesis includes several more proteins (Figure 2.3), involved in a complex network of interactions. Understanding the control of angiogenesis and the roles of regulators (positive and negative) as well as all the molecules that can trigger an angiogenic response is of paramount importance. Advances made through research have greatly aided towards this purpose and future advances can significantly improve the understanding and targeting of angiogenesis-related diseases.

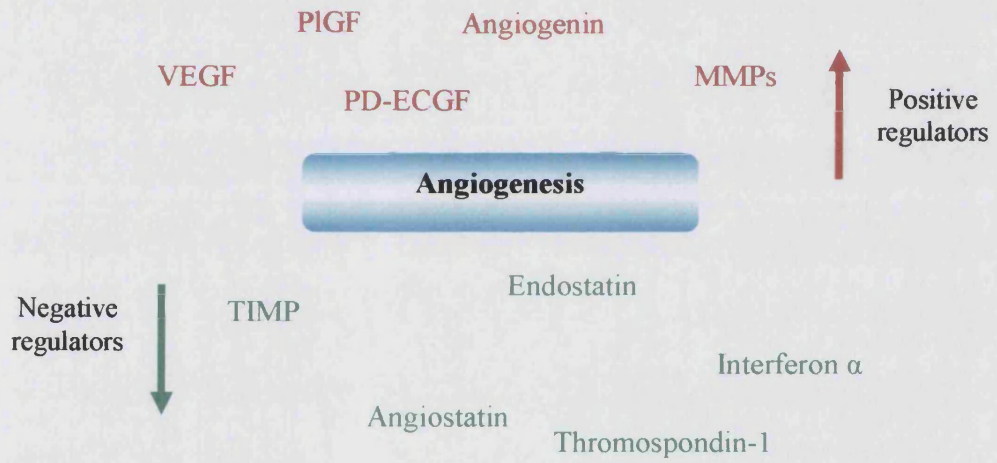


Figure 2.3: Angiogenesis regulation depends on the balance between angiogenic (positive regulators) and angiostatic molecules (negative regulators); imbalance in either positive or negative regulators can lead to excessive or insufficient angiogenesis.

Several enzymes that are involved in angiogenesis have been studied and characterised to date. PD-ECGF/TP has been the main focus of my PhD study. Crystallographic and molecular biology techniques have been applied to acquire more information about this enzyme and better understand its structure and function; see Chapter III for details.

Due to the number of diseases linked to excessive angiogenesis, its negative regulators have attracted a lot of interest. One such regulator, TIMP from *Drosophila melanogaster*, *d*TIMP, has been studied during my PhD. Modeling work on *d*TIMP, a homologue of *h*TIMP (human TIMP), provides insights in the evolutionary relationship of TIMPs from different organisms, (see Chapter IV for details) especially since functional similarity does not always imply structural similarity.

Chapter – III –

Human Thymidine Phosphorylase

3.1 Introduction to nucleoside phosphorylases

Nucleoside phosphorylases catalyse the reversible phosphorolysis of purine and pyrimidine nucleosides, a reaction of key importance for the nucleotide salvage pathway. The phosphorolysis of purine and pyrimidine nucleosides occurs by the cleavage of the C-N glycosidic bond by a phosphate ion to yield a free base and 2-deoxy-D-ribose-1-phosphate. Two families of nucleoside phosphorylases have been described: purine phosphorylases (PNPs), which cleave the glycosidic bond of purines (Figure 3.1) and pyrimidine phosphorylases (PyNPs), which cleave the glycosidic bond of pyrimidines (Figure 3.2), while no phosphorylase has been described to cleave the glycosidic bond of cytidine (Levene and Medigreceanu, 1911).

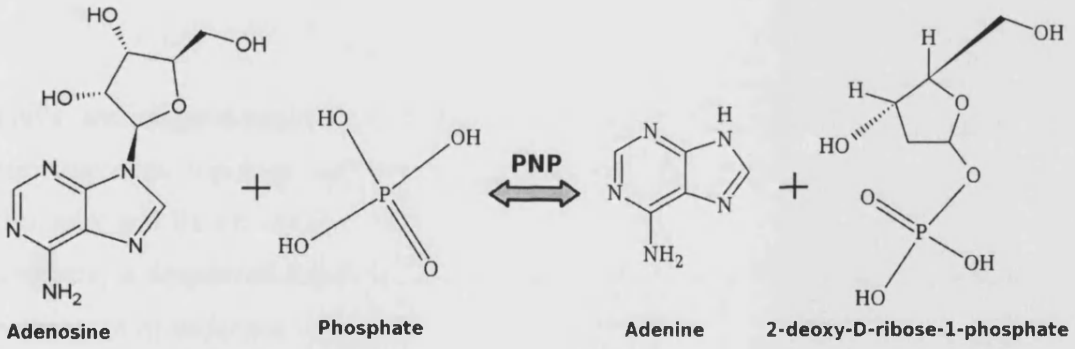


Figure 3.1: Schematic representation of the reaction catalysed by adenosine phosphorylase, a member of the purine nucleoside phosphorylase superfamily (PNP); adenosine and phosphate are used as substrates (for the catabolic reaction) while adenine and 2-deoxy-D-ribose-1-phosphate are the products of the reaction.

Thymidine Phosphate Thymine 2-deoxy-D-ribose-1-

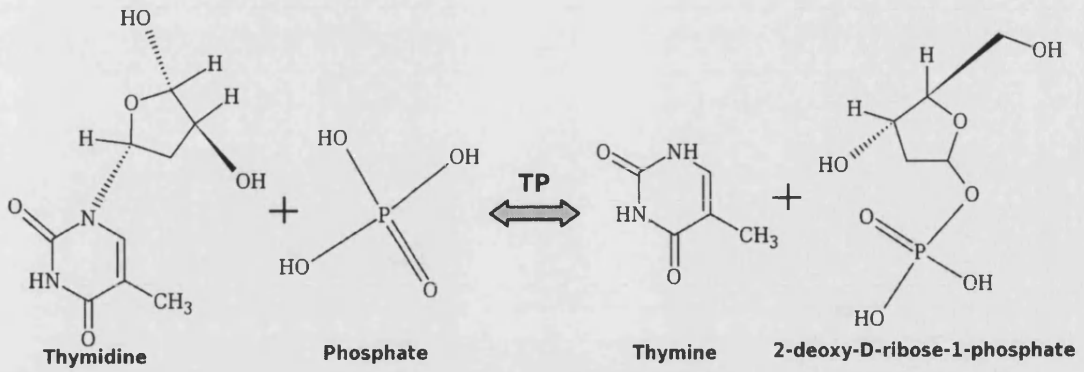


Figure 3.2: Schematic representation of the (catabolic) reaction catalysed by thymidine phosphorylase, a member of the pyrimidine nucleoside phosphorylase superfamily (PyNP). Thymidine and phosphate are used as substrates while thymine and 2-deoxy-D-ribose-1-phosphate are the products of the reaction.

PNPs are single-domain molecules, with either a trimeric or hexameric quaternary structure; this topology appears to be conserved among all members of the family (Pugmire and Ealick, 2002). PyNPs are two-domain molecules with a dimeric quaternary structure, a conserved topology amongst the members of the family and a significant percentage of sequence identity (>30%) (Walter *et al.*, 1990; Pugmire and Ealick, 1998; Pugmire *et al.*, 1998). In lower organisms PyNPs accept both thymidine and uridine, while in mammals and other higher organisms there are two types which distinguish between the two: one specific for uridine and one specific for thymidine (Paege and Schlenk, 1952). The differences in the fold between PNPs and PyNPs suggest that the two families have evolved separately despite the similarity of the reactions they catalyse.

3.2 Thymidine phosphorylase (TP) and platelet-derived endothelial cell growth factor (PD-ECGF)

Mammalian thymidine phosphorylase (TP) and its role in the nucleotide salvage pathway was first described by Friedkin and Roberts in 1954 (Friedkin and Roberts, 1954). Biochemical characterisation of TP demonstrated that it has low substrate specificity (Desgranges *et al.*, 1981) and readily catalyses the reversible phosphorolysis (having primarily a catabolic function; Iltzsch *et al.*, 1985) of thymidine and 2'-deoxyuridine to their respective base and to 2-deoxy-D-ribose-1-phosphate, as well as that of some pyrimidine analogues (Piper and Fox, 1982; Desgranges *et al.*, 1983; Verri *et al.*, 2000). TP accepts and acts only on naturally occurring D-thymidine and D-thymidine analogs but not on their L-counterparts, signifying the enantioselectivity of the enzyme (Spadari *et al.*, 1995).

In 1987 a protein found to induce the [^{3}H] thymidine incorporation, endothelial cell migration *in vitro* and angiogenesis *in vivo* was isolated from human platelets (Miyazono *et al.*, 1987; Ishikawa *et al.*, 1989) and was thus given the name platelet-derived endothelial cell growth factor (PD-ECGF). Purified PD-ECGF is a single chain polypeptide which appears to act as a homodimer of ~90kD molecular weight (Ishikawa *et al.*, 1989). The gene for PD-ECGF was later identified (Hagiwara *et al.*, 1991) to consist of 10 exons and to be located on chromosome 22, dispersed over a 4.3kb region (Figure 3.3).

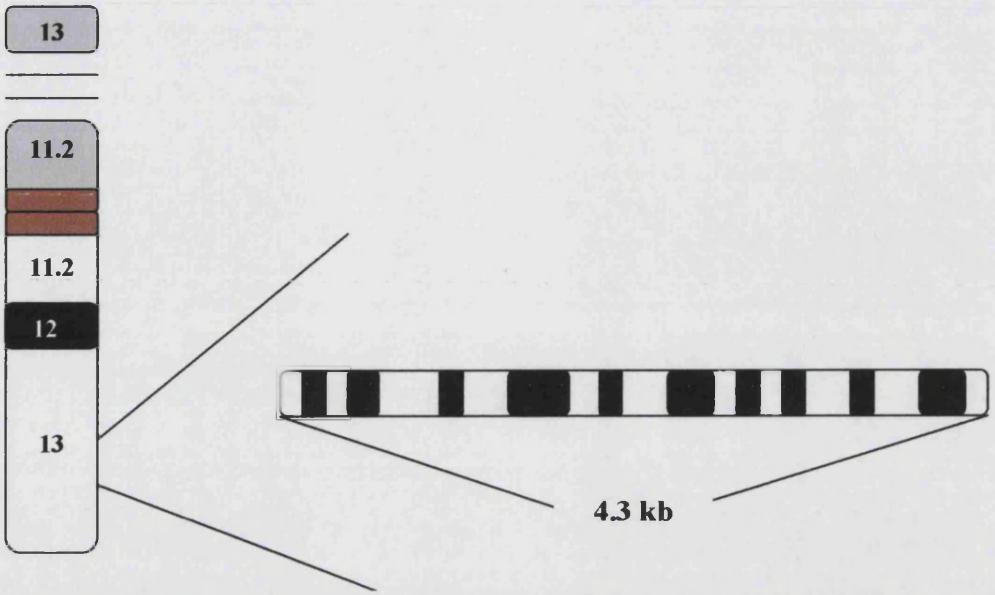


Figure 3.3: Schematic representation of chromosome 22 and the gene encoding for TP. Ten exons (represented by black boxes) are present in the *TP* gene, which is dispersed over a 4.3kb region. (Figure adapted from data produced by the Chromosome 22 Group at the Sanger Institute; Collins *et al.*, 2003).

Interestingly the promoter region of human PD-ECGF seems to lack characteristic features of eukaryotic promoters, such as a 'TATA' box or a 'CCAAT' box (Miyazono *et al.*, 1991). Post-translational modification and glycosylation do not seem to take place in the case of PD-ECGF; the enzyme stays inside the cell and is secreted very slowly without any further processing, making it a 'non-classical' secretory protein (Usuki *et al.*, 1989). The only post-translational modification that the enzyme seems to undergo is the phosphorylation of serine residues; this could be a way to facilitate its secretion, otherwise impossible due to the lack of a signal peptide (Asai *et al.*, 1992b). Amino acid sequence analysis and gel chromatography revealed that PD-ECGF is identical to TP (Usuki *et al.*, 1992). In 1992 Asai and co-workers identified PD-ECGF/TP to be identical to gliostatin (GLS), a glial growth inhibitory factor from human neurofibroma (Asai *et al.*, 1992).

3.3 PD-ECGF/TP in physiological angiogenesis

Angiogenesis is the generation of new capillary blood vessels from pre-existing vasculature. It occurs in several physiological and pathological processes (see Chapter II). As a multi-step process it involves a wide number of mediators, positive (up-regulation) and negative (down-regulation). The implication of TP in angiogenesis has been demonstrated by its ability to stimulate endothelial cell migration *in vitro* (Miyazono *et al.*, 1989) and angiogenesis *in vivo*, in models such as the chick chorioallantoic membrane (CAM) and in MCF-7 breast cancer cells, where TP induced an increase in the levels of tumour vascularisation (Haraguchi *et al.*, 1994; Moghaddam *et al.*, 1995; Liekens *et al.*, 2002).

In contrast to other angiogenic factors which bind to cell surface receptors and activate signal transduction pathways, TP has not been shown to recognise and bind any. Enzymatic activity assays on TP where active site residues have been mutated, demonstrated that its catalytic activity is indispensable for its angiogenic activity; limiting the enzymatic activity of TP results in abolishing the angiogenic activity of the enzyme (Haraguchi *et al.*, 1994; Moghaddam *et al.*, 1995; Miyadera *et al.*, 1995).

One of the key roles of TP is the maintenance of a balance in the nucleotide pool and control of nucleic acid homeostasis. This is achieved by retaining a constant and correct supply of deoxyribonucleoside triphosphates (dNTPs) for DNA repair and replication (Figure 3.4). It is important to note that inadequate supply of dNTPs can be lethal while excess can lead to mutagenesis (Kornberg and Baker, 1991; Hübscher and Spadari, 1994).

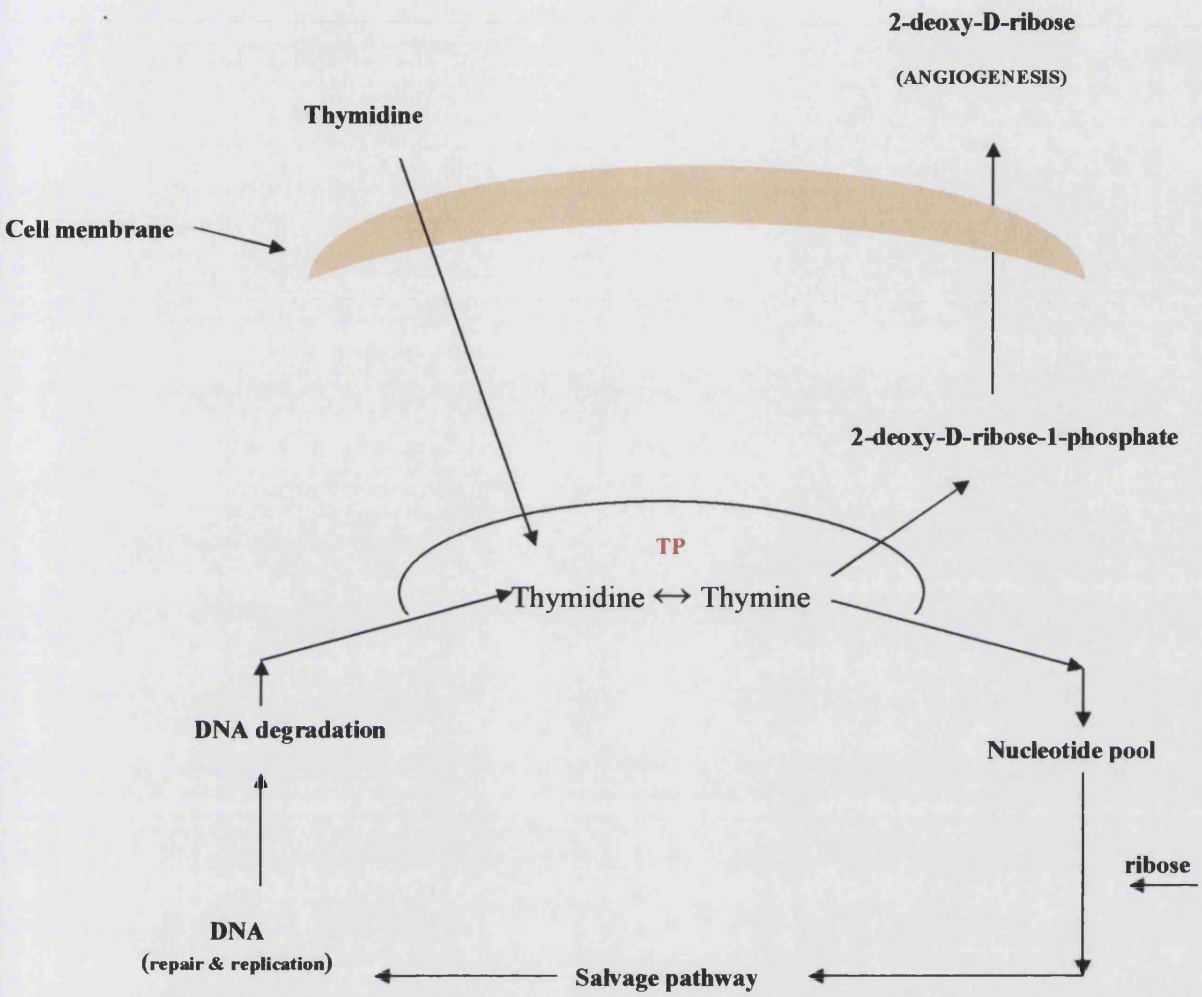


Figure 3.4: The role of TP in nucleic acid homeostasis. TP catalyses the phosphorolysis of thymidine to thymine and 2-deoxy-D-ribose-1-phosphate. Of the two products, the latter is dephosphorylated in the cytoplasm and can leave the cell; 2-deoxy-D-ribose has also been shown to have angiogenic activity. Cytostatic thymidine enters the cell in a non-energy dependent manner and is cleaved by TP, driving in this way the salvage pathway and maintaining a constant supply of nucleotides for DNA repair and replication. Imbalances in this cycle affect both nucleic acid homeostasis and angiogenic activity of TP, possibly mediated by 2-deoxy-D-ribose. (Figure adapted from: Brown and Bicknell, 1998).

One of the products of the reaction catalysed by TP is 2-deoxy-D-ribose-1-phosphate (2dR1P), proposed to be responsible for the angiogenic activity of the enzyme. More specifically 2dR1P, is dephosphorylated in the cytoplasm to 2-deoxy-D-ribose, which is cell permeant and can thus leave the cell (Figure 3.4). This monosaccharide, which is an endothelial cell chemoattractant has also been shown to induce angiogenesis (Haraguchi *et al.*, 1994; Moghaddam *et al.*, 1995; Miyadera *et al.*, 1995), probably in a manner similar to that of glucose, which by acting as an energy source attracts endothelial cells along its concentration gradient (Vogel *et al.*, 1993).

In contrast to other chemoattractants that bind to specific endothelial cell surface receptors, 2-deoxy-D-ribose has not been shown to bind any. The exact mechanism by which TP and 2-deoxy-ribose induce angiogenesis is not yet completely understood; Brown and co-workers proposed a hypothesis regarding the implication of 2-deoxy-D-ribose in angiogenesis (Brown *et al.*, 2000). This hypothesis was based on the nature of 2-deoxy-D-ribose as a strongly reducing sugar; its presence can generate oxygen radicals that induce oxidative stress response angiogenic factors, like VEGF, MMPs and IL-8.

3.4 PD-ECGF/TP in pathological angiogenesis

TP has been shown to play a crucial role in pathological angiogenesis (Figure 3.5). Cases of carcinomas and excessive angiogenesis where TP was found to be overexpressed have been widely reported while TP deficiency has also been related to certain clinical conditions.

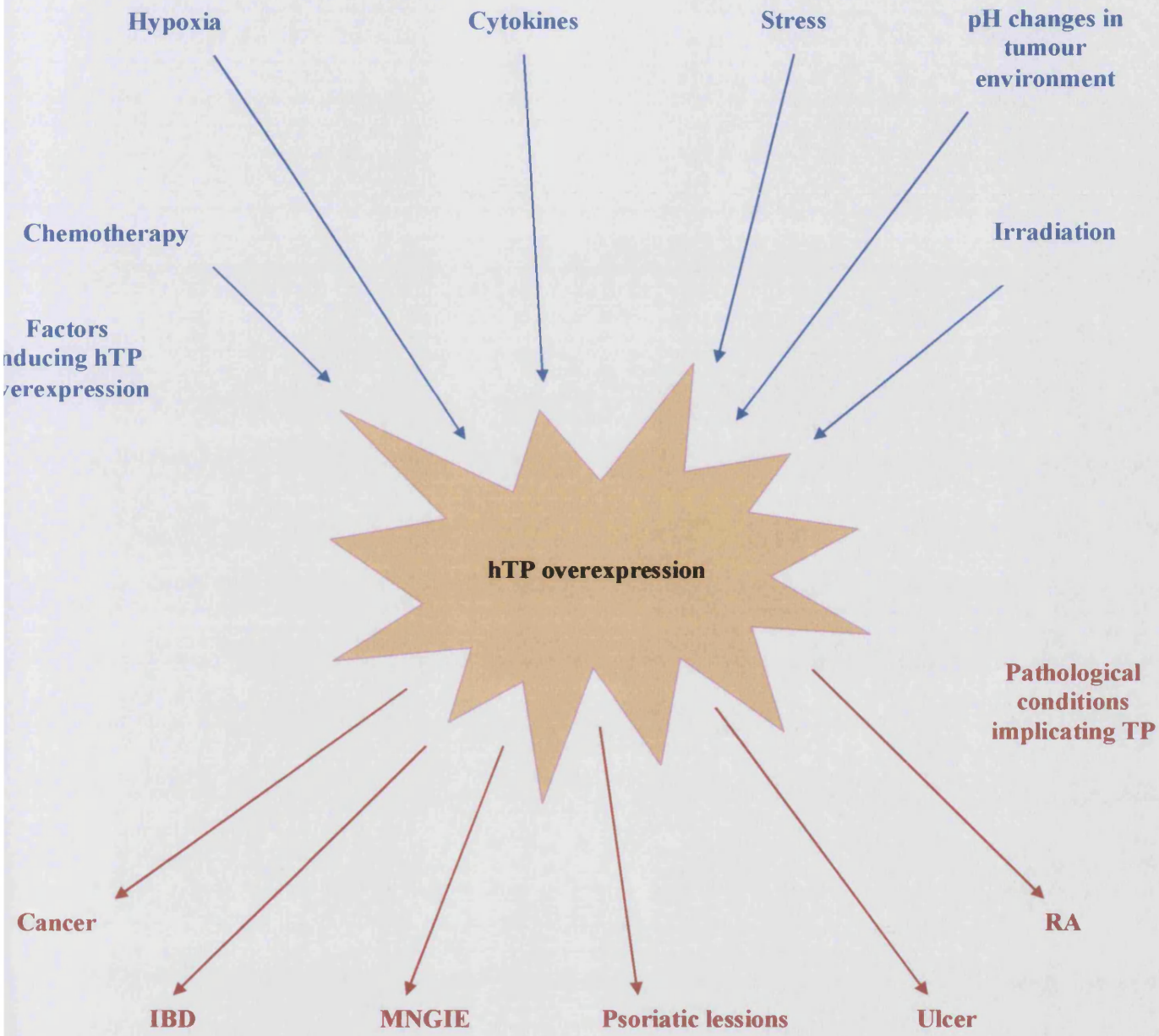


Figure 3.5: Factors that induce TP upregulation include hypoxia, cytokines, chemotherapeutic drugs, stress, changes in tumour environment and irradiation. Overexpression of TP has been observed in various pathological conditions, such as cancer, ulcers, rheumatoid arthritis, psoriatic lesions, irritable bowel disease and MNGIE.

3.4.1. PD-ECGF/TP and cancer

TP has been shown to overexpress in various types of tumours (Figure 3.5); histological analyses have revealed increased levels of TP in tumours compared to those of non-neoplastic regions of the same organs. Due to its angiogenic activity, increase in TP expression leads to increase in the microvasculature and has been observed in cases such as colorectal, gastric, breast, pancreatic, cervical, renal, prostate and endometrial tumours (Reynolds *et al.*, 1994; Toi *et al.*, 1995; Takebayashi *et al.*, 1996; Imazano *et al.*, 1997; Takao *et al.*, 1998; Yao *et al.*, 2001; Sivridis *et al.*, 2002). Additionally carcinomas that were TP-positive were linked to poorer prognosis compared to TP-negative carcinomas, implicating thus TP in the development of malignancy in tumours (Akiyama *et al.*, 2004). Hence, increase in tumour growth, poorer prognosis and resistance to apoptosis are three key processes in which TP has been shown to be involved, mostly due to its angiogenic activity.

Inflammatory mediators associated with physical stress have also been shown to upregulate TP (Figure 3.5). More specifically, even in early stages of non-invasive cancers, hypoxic environments are developed which are rich in cytokines. Cytokines such as tumour necrosis factor (TNF) α and interleukin 1 can alter the phenotype of neighbouring stromal cells and have been shown to upregulate the expression of TP in malignant cells (Eda *et al.*, 1993; Goto *et al.*, 2001; Zhu *et al.*, 2003). Upregulation of TP in conditions of environmental stress and inflammation has been confirmed by the stimulation of its expression in hypoxia or hypoglycaemia (Griffiths *et al.*, 1997).

TP has been shown to co-express with other angiogenic molecules, such as VEGF, an important mediator of vascular development and growth. This suggests that the two molecules may act cooperatively in neovascularisation. One of the products of the catabolic reaction catalysed by TP, 2-deoxy-D-ribose has also been implicated in neovascularisation of tumours as it has been shown to stimulate cell migration in the endothelium, necessary for the development of new capillaries and maintenance of oxygen and nutrients for tumour growth, progression and metastasis (Zhang *et al.*, 1995).

3.4.2. PD-ECGF/TP in cancer treatment

Overexpression of TP has also been shown to be induced by several types of cancer treatment. Chemotherapeutic drugs, such as taxanes (Sawada *et al.*, 1998; Endo *et al.*, 1999) and cyclophosphamide (Sawada *et al.*, 1998; Endo *et al.*, 1999) as well as irradiation have proved to induce expression of TP in various solid tumours (Figure 3.5). This is probably due to the induction of its expression *via* cytokines (e.g. TNF-α) which are upregulated by the use of chemotherapeutic drugs. Treatments which aim at altering the tumour environment, either by changes in the pH or by inducing hypoxia have also been shown to induce TP. Hence, combinatorial treatments that aim at inactivating TP and target tumour growth and progression are considered to be essential for cancer treatment.

3.4.3. PD-ECGF/TP and protumour stroma

Studies on tumour stromal cells revealed TP overexpression in tumour-associated macrophages as well as in stromal cells, suggesting TP is implicated in poorer prognosis in cases such as breast cancer, astrocytic tumours and uterine endometrial cancers (Toi *et al.*, 1995b; Seki *et al.*, 1999; Yao *et al.*, 2001). The implication of TP in such cases is considered to be mediated by its metabolite 2-deoxy-D-ribose-1-phosphate, which is assumed to change the property of stromal cells from non-protumour or antitumour to protumour, *via* the stimulation of angiogenic, inflammatory, tissue digestive and steroidogenic molecules (Brown *et al.*, 2000). Based on these findings, Toi and co-workers proposed the conversion of TP-positive to TP-negative stroma as a treatment for cancer, suggesting that protumour stroma can change back to antitumour stroma (Toi *et al.*, 2005).

3.4.4. PD-ECGF/TP and apoptosis

Implication of TP in many solid tumours is not only based on its angiogenic properties that stimulate neovascularisation; TP has also been shown to mediate inhibition of tumour cell apoptosis. Hypoxia (lack of oxygen) enhances apoptosis (cell death), which in turn induces the expression of angiogenic factors, such as VEGF and TP (Figure 3.4; Sheweiki *et al.*, 1992; Yao *et al.*, 1995; Graeber *et al.*, 1996). During tumour growth it is

common for the centre of the tumour to become hypoxic, leading to apoptosis and necrosis. This is however usually limited to the centre of the tumour, suggesting that there are specific mechanisms recruited by the remaining tumour cells in order to maintain tumour progression. TP has been shown to be involved in one of these mechanisms: it appears to provide resistance to apoptosis when the latter is induced by conditions such as hypoxia (Kitazono *et al.*, 1998). The enzymatic activity of TP did not seem to be essential for this role although its metabolite, 2-deoxy-D-ribose, has also been shown to prevent hypoxia-induced cell apoptosis, indicating that both enzyme and metabolites are cytoprotective. Due to its angiogenic activity and cooperation with other angiogenic molecules, TP appears to be crucial for the escape of tumour cells from apoptotic signals, including physical stress and anticancer treatment (Toi *et al.*, 2005).

3.4.5. PD-ECGF/TP and VEGF

In several cases of highly vascularised tumours, such as head, neck, breast, lung, colorectal, cervical and endometrial tumours, TP has been found to co-express with VEGF (Toi *et al.*, 1995a; Fukuiwa *et al.*, 1999; O'Byrne *et al.*, 2000; Yao *et al.*, 2001; Kabuubi *et al.*, 2006). The presence of an SP1 transcription factor binding site in the promoter regions of both TP and VEGF explains the similar transcription regulation of *TP* and *VEGF* genes (Hagiwara *et al.*, 1991), hence the cooperative action of VEGF and TP. Experiments in stage I lung adenocarcinoma revealed that TP induction is mostly noticed on the tumour stroma while VEGF induction is most clearly seen in the tumour cells (Kojima *et al.*, 2002); it can be thus assumed that cooperation of TP and VEGF may result in a more aggressive tumour phenotype.

3.4.6. PD-ECGF/TP in inflammatory diseases

Apart from its implication in various types of tumours TP has also been shown to be involved in various inflammatory and angiogenic diseases (Figure 3.5). Hammerberg and co-workers (Hammerberg *et al.*, 1991) have identified TP overexpression in psoriatic lesions. Higher TP concentration has also been detected in sections of colonic mucosa of inflammatory bowel disease (IBD), predominately in macrophages and fibroblasts, while increase in microvessel density was observed in the inflamed mucosa (Saito *et al.*, 2003).

Kasugai and co-workers identified its implication in gastric ulcers (Figure 3.5). The human plasma concentration of GLS/PD-ECGF/TP was significantly higher in patients with intractable gastric ulcer. The tissue content was significantly higher at the gastric ulcer edge than in either the fundic or pyloric region. GLS/PD-ECGF infusion delayed ulcer healing in a dose-dependent manner (Kasugai *et al.*, 1997). These results suggest that gastric tissue and/or circulating GLS/PD-ECGF are part of the pathology of gastric ulcers.

Gliostatin, previously identified to be identical to PD-ECGF/TP, has been implicated in rheumatoid arthritis (RA; Figure 3.5; Asai *et al.*, 1993; Takeuchi *et al.*, 1994). Presence of TP in synovial fluid was considered to be indicative of neovascularisation in rheumatoid synovial tissues (Waguri *et al.*, 1997). Implication of TP in RA is proposed to be *via* the production of cytokines such as TNF-α and IL-1 from fibroblast-like synoviocytes, which in turn stimulate the production and secretion of TP (Waguri *et al.*, 1997). The ability of TP to induce MMP production would further explain the extensive joint destruction observed in RA (Muro *et al.*, 1999; Ieda *et al.*, 2001).

The gene encoding for *TP* is located on chromosome 22 (see *3.2*). Mutation on the origin of this gene has been shown to cause an autosomal recessive disease, known as mitochondrial neurogastrointestinal encephalomyopathy (MNGIE). Deficiency in TP activity leads to increased levels of thymidine and deoxyuridine, as these are not catalysed. Imbalances caused by elevated thymidine concentrations can alter cell growth and viability (Barclay *et al.*, 1982), by inducing deletions and point mutations in mitochondrial DNA (Spinazzola *et al.*, 2002). Further research on TP and its association with MNGIE revealed that the mutation in the origin of the gene noticed in MNGIE cases is also present in some individuals with normal TP activity, while overlapping on the same chromosome was found to be a gene encoding for SCO2 (Haraguchi *et al.*, 2002). SCO2 is a COX assembly gene (cytochrome *c* oxidase), located next to human *TP* gene (exon 10), which has been previously reported in patients with foetal infantile cardioencephalomyopathy and has also been shown to be associated with severe COX deficiency in heart and skeletal muscle (Haraguchi *et al.*, 2002).

It is thus evident that the wide number and severity of diseases in which TP has been shown to be involved (either directly or *via* its metabolites) make the need for suitable and effective inhibitors of its enzymatic activity imperative.

3.4.7. Inhibitors of PD-ECGF/TP

As TP is involved in various inflammatory, angiogenic and cancer conditions inhibition of its activity has been the target of several research groups. Even when the target is the angiogenic activity of TP, the enzymatic activity has to be abolished as the first is dependent on the latter. Several types of inhibitors have been previously designed and tested for their inhibitory activity. An example is 7-Deazaxanthine (7DX), the first purine derivative, which inhibited thymidine phosphorylase in a concentration dependent manner (Balzarini *et al.*, 1998). Using 7DX as a basis and covalently linking it to an alkyl phosphonate moiety, lead to the design of a new inhibitor; TP65. The latter inhibitor, (Liekens *et al.*, 2002), was later proposed to inhibit thymidine phosphorylase by keeping it in an inactive conformation. Another type of inhibitor designed for thymidine phosphorylase had a homopthalmide skeleton. A potent inhibitor of this class was 2-(2,6-diethyl)-7-nitro-1,2,3,4-tetrahydroisoquinoline-1,3-dione, which was shown to cause a mixed-type inhibition of thymidine phosphorylase (Kita *et al.*, 2001). Fukushima *et al.*, in 2000, reported one of the most potent inhibitors designed for thymidine phosphorylase, (5-chloro-6-[1-(2-iminopyrrolidinyl)methyl] uracil hydrochloride), TPI. TPI was described to have an IC_{50} value of 35nM (Fukushima *et al.*, 2000) and it was suggested to suppress tumour growth by increasing the apoptotic rate. Pyrimidine derivatives have always been considered to be the most popular targets for drug design and development. Murray and co-workers (Murray *et al.*, 2002) reported a series of bicyclic pyridinium-substituted uracil and thymine analogues that have been designed for thymidine phosphorylase. Their efficacy was tested on *E.coli* TP: the 6-amino-5-substituted derivatives have been shown to be potent inhibitors of TP, the most potent being 6-amino-5-bromouracil, with and IC_{50} of 1.1mM. Brown and Bicknell (1998) reported the efficacy of 5-nitrouracil in TP inhibition; IC_{50} of 414 μM. Reigan (Reigan *et al.*, 2005) described the design and development of a class of inhibitors that act as prodrugs and rely on the tumour environment (in particular the hypoxic conditions and high concentration

of xanthine oxidase present in many tumours) for their conversion. One such example is the design of 5-halo-6'-[(2'-nitroimidazol-1'-yl)methyl]uracil, reduced to 2'-amino analogues by xanthine and reported to be a potent inhibitor of TP (Reigan *et al.*, 2005).

An interesting type of inhibition is that of the metabolite of TP, 2-deoxy-D-ribose by its enantiomer 2-deoxy-L-ribose. The latter has been shown to inhibit endothelial cell migration occurring as a response to 2-deoxy-D-ribose (Akiyama *et al.*, 2004; Nakajima *et al.*, 2004). The stimulation of other angiogenic molecules by 2-deoxy-D-ribose was also reduced in the presence of its enantiomer, indicating its potency in inhibiting this angiogenic mediator. Although 2-deoxy-L-ribose has been shown to reduce the angiogenic activity of TP, it does not seem to affect its enzymatic activity; this is of key importance as maintenance of nucleic acid homeostasis is crucial.

Most inhibitors that are currently designed and tested are prodrugs that require TP for their conversion. One such example is doxifluridine, an oral prodrug, which upon conversion from TP becomes 5-fluorouracil (5FU). This however is a rate-limiting drug as toxicity is accumulated in areas where TP is expressed; not only in tumours but in the gastrointestinal tract as well. A new drug, currently in clinical trials, capecitabine, was designed so as to pass through the gastrointestinal tract at its intact form and deliver the converted 5FU only to cancer cells (Miwa *et al.*, 1998).

For a successful inhibitor, parameters such as the nature of the inhibitor, the size and the number of interactions with the enzyme have to be considered. In the present study hTP has been examined at its native state and in complex with a pyrimidine based analogue, 5-iodouracil, 5IUR, which greatly resembles the chemotherapeutic agent 5-fluorouracil (Brown and Bicknell, 1998).

3.5 Pyrimidine nucleoside phosphorylases : an overview

The family of pyrimidine nucleoside phosphorylases consists of thymidine and uridine phosphorylases (see *3.1*). The similarity of the substrates they recognise (Figure 3.6) as well as the reaction they catalyse would probably lead to the assumption that the two

types of pyrimidine nucleoside phosphorylases are also similar in structure; primary to quaternary. Interestingly this is not the case: uridine phosphorylase differs greatly from thymidine phosphorylase, making the members of the PyNP superfamily even more interesting.

Uridine **Thymidine**

Figure 3.6: Structure of uridine and thymidine, the two substrates catalysed by uridine nucleoside phosphorylase and thymidine nucleoside phosphorylase respectively.

3.5.1. Uridine nucleoside phosphorylase

Uridine phosphorylase (UP; Pugmire and Ealick, 2002) is an enzyme of key role in the pyrimidine salvage pathway: it catalyses the reversible phosphorolysis of uridine to uracil and ribose-1-phosphate. As previously described (see *3.1*) two typical folds represent members of the nucleoside phosphorylase superfamily: a fold which is considered to be characteristic of PNPs and a different one, characteristic of PyNPs. Although functionally UP is a member of the PyNP superfamily, structurally it appears to differ greatly from the typical fold expected for members of this family. The fold it acquires is a single subunit α/β-fold (Figure 3.7).

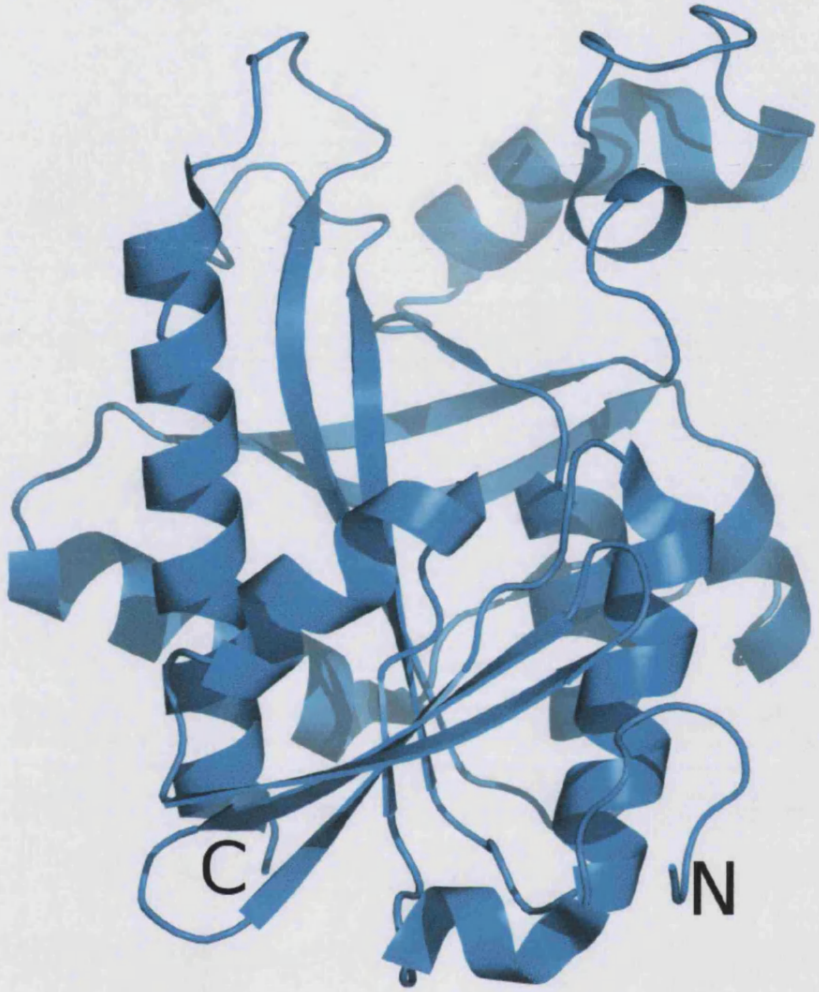


Figure 3.7: Cartoon representation of UP monomer from *Escherichia coli* (PDB ID 1RXY; Caradoc-Davies *et al.*, 2004). Uridine phosphorylase is a one-domain molecule (a single subunit), that acquires an α/β-fold.

The quaternary structure of the enzyme consists of trimers or hexamers. One such example is the structure of *E.coli* UP (Burling *et al.*, 2003; Caradoc-Davies *et al.*, 2004) which acquires a hexameric quaternary structure (Figure 3.8).

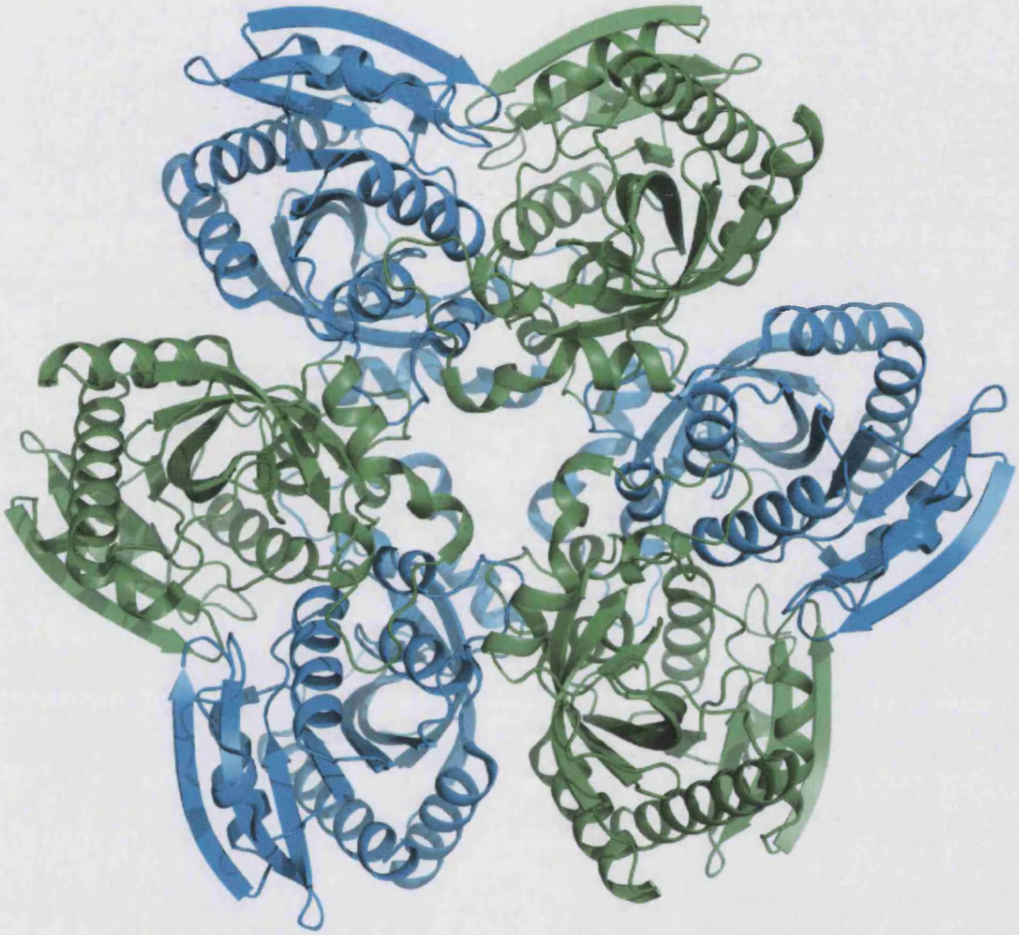


Figure 3.8: Cartoon representation of the quaternary structure of UP from *Escherichia coli* (PDB ID 1RXY; Caradoc-Davies *et al.*, 2004). The hexameric quaternary structure observed for UP (PyNP) is typically observed for members of the PNP superfamily.

Three binding sites have been identified in *E.coli* UP: a phosphate binding site, a ribose binding site and a uracil binding site. The phosphate binding site appears to consist of almost identical residues to that of *E.coli* PNP, probably due to the evolutionary relationship and conserved function of these enzymes. The ribose binding site consists of four key residues in both *E.coli* UP and *E.coli* PNP, suggesting that the same overall

pattern of hydrogen bonding to ribose exists in both cases (Caradoc-Davies *et al.*, 2004). Comparison of the sequence of UP from *E.coli* and human revealed that residues involved in ribose and uracil binding are conserved and probably involved in the same network of interactions as that of *E.coli* PNP. These findings suggest that UP and hexameric PNP diverged from a common ancestor (Pugmire and Ealick 2002), while the few differences observed between the two structures are probably the evolutionary event that separated UP from the ancestral form of hexameric PNPs.

Structural studies on native and complex forms of *E.coli* UP (Caradoc-Davies *et al.*, 2004) revealed that it displays induced fit upon substrate binding (specific for pyrimidines) and acquires a closed conformation, proposing thus a type of 'domain movement', essential for catalysis to occur. Regions that appear to be disordered in the native form of the enzyme, show significant improvement in the closed, bound state.

Although the sequence, structure and mechanism of *E.coli* UP was proposed to resemble greatly those of PNPs, the enzyme appears to be specific only for pyrimidines (Vita *et al.*, 1986). These findings present an interesting case of common ancestry, evolutionary divergence and structure-function relationships of enzymes with key roles in the nucleoside salvage pathway.

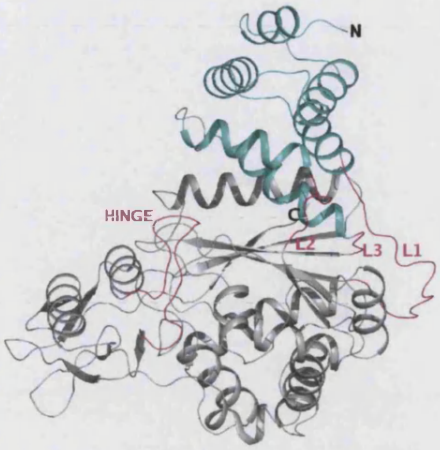
3.5.2. Thymidine nucleoside phosphorylase structures from Escherichia coli, Bacillus stearothermophillus and Human species.

A wealth of structural information about thymidine nucleoside phosphorylase has come from studies on *E.coli* and *B. stearothermophillus* TP. The structure of *E.coli* thymidine phosphorylase (EcTP), the first structure of TP to be determined, was first reported by Walter and co-workers in 1990 (Figure 3.9; Walter *et al.*, 1990) at 2.8Å resolution, with a sulphate ion bound on the phosphate binding site. The structure of *Bacillus stearothermophillus* PyNP (BsPyNP) was later reported by Pugmire and Ealick in 1998 (Figure 3.9; Pugmire and Ealick, 1998) in complex with pseudouridine and phosphate. The structure of human thymidine phosphorylase (hTP; Figure 3.9) was first described by Norman and co-workers in 2004 (Norman *et al.*, 2004) in complex with a small molecule

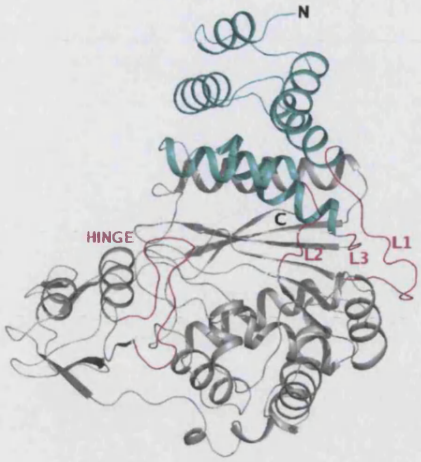
inhibitor, TPI. Another complex of hTP was later determined by El Omari, where thymine was identified in the active site of the enzyme, although thymidine was present in the crystallisation conditions (El Omari *et al.*, 2006). While *E.coli* and human TP have been described to be specific for 2'-deoxy-ribosides, BsPyNP showed no such specificity.

The first thymidine phosphorylase to be structurally characterised, EcTP, was identified to consist of two domains, a small α-helical domain, composed of six helices and shown to be involved in interactions at the dimer interface. The larger mixed α/β domain was identified to consist of a central six-stranded mixed β-sheet, surrounded by two helices and a smaller, four-stranded antiparallel β-sheet, with no evident interactions with the central six-stranded mixed β-sheet (Figure 3.9).

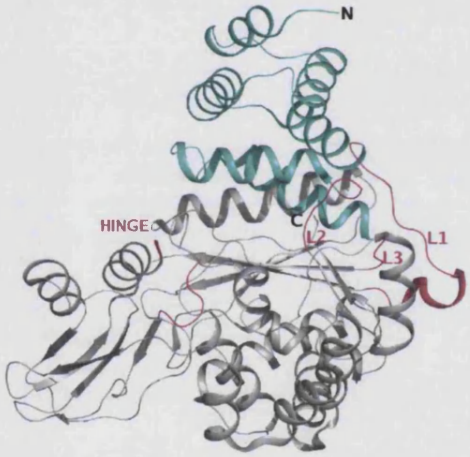
Bacterial TP is smaller than its mammalian homologue: each monomer has a molecular weight of 45kDa, composing a homodimer of around 90kDa, while each monomer of mammalian TP has a molecular weight of around 50kDa, with a dimer molecular weight of around 100kDa. On the sequence level bacterial TP shares ~41% sequence similarity with human TP (Walter *et al.*, 1990; Brodsky *et al.*, 1992). Structurally, bacterial TP resembles greatly its mammalian homologue: they have the characteristic PyNP fold of a small α and larger α/β domain, 'held' together by three polypeptide loops: L1, L2 and L3 (Figure 3.9; Walter *et al.*, 1990; Norman *et al.*, 2004; El Omari *et al.*, 2006). Structural characterisation of TP from bacterial sources initiated the design and development of drugs targeting these enzymes, while mechanisms of TP activity, proposed on the basis of structural findings, have 'directed' the design and development of drugs targeting its enzymatic activity and, subsequently, its angiogenic activity.



EcTP
(Walter *et al.*, 1990)



BsPyNP
(Pugmire and Ealick, 1998)



hTP
(Norman *et al.*, 2004)

Figure 3.9: Cartoon representation of the secondary structure of the three thymidine phosphorylases; EcTP, BsPyNP and hTP. Shown in cyan is the α domain and in grey is the mixed α/β domain. Highlighted in pink are the three loop regions (L1, L2 and L3) and the hinge region (HINGE), proposed to allow for flexibility, necessary for domain movement upon occupation of the phosphate and substrate binding sites. In the structure of hTP (first described by Norman *et al.*, 2004) the hinge region was proteolytically cleaved.

3.5.3. Binding sites of TP

The active site where the substrate (thymidine) binds is proposed to be located at the interface between the α domain and the mixed α/β domain (Figure 3.10; Walter *et al.*, 1990; Pugmire and Ealick, 1998). A number of charged residues form a network of interactions with molecules bound at the active site, stabilising them in the active site cleft (Walter *et al.*, 1990; Pugmire and Ealick, 1998; Norman *et al.*, 2004). Residues identified to interact with molecules bound at the active site of EcTP, BsPyNP and hTP appear to be conserved amongst species.

A phosphate binding site was identified to be located at the mixed α/β domain, access to which was proposed to be *via* the cavity between the two domains (Figure 3.10; Walter *et al.*, 1990; Pugmire *et al.*, 1998a; Pugmire and Ealick, 1998). The residues previously proposed to be essential for phosphate binding have also been identified in hTP, suggesting that these residues are conserved among species. A glycine-rich loop has been proposed to be involved in the binding of the phosphate ion. Similar glycine-rich loops have been identified in other nucleotide-binding proteins (Dreusicke and Schulz, 1986). Due to its high content in glycine residues it is considered to be very flexible and aid the movement of the two domains (Pugmire *et al.*, 1998a). This glycine-rich loop region is conserved in all known TP structures, signifying its potential functional implication.

Two more 'pockets' were identified in the structure of BsPyNP: one potential metal ion (most probably Ca^{2+} or K^{+}) binding site and one hydrophobic pocket. The hydrophobic pocket surrounds the 5'-position of the pyrimidine ring bound on the active site, as seen in this crystal structure (Pugmire and Ealick, 1998). The same hydrophobic pocket was later identified in hTP; in the structure of hTP-TPI this pocket appears to hold the chlorine atom of TPI (hTP; Norman *et al.*, 2004).

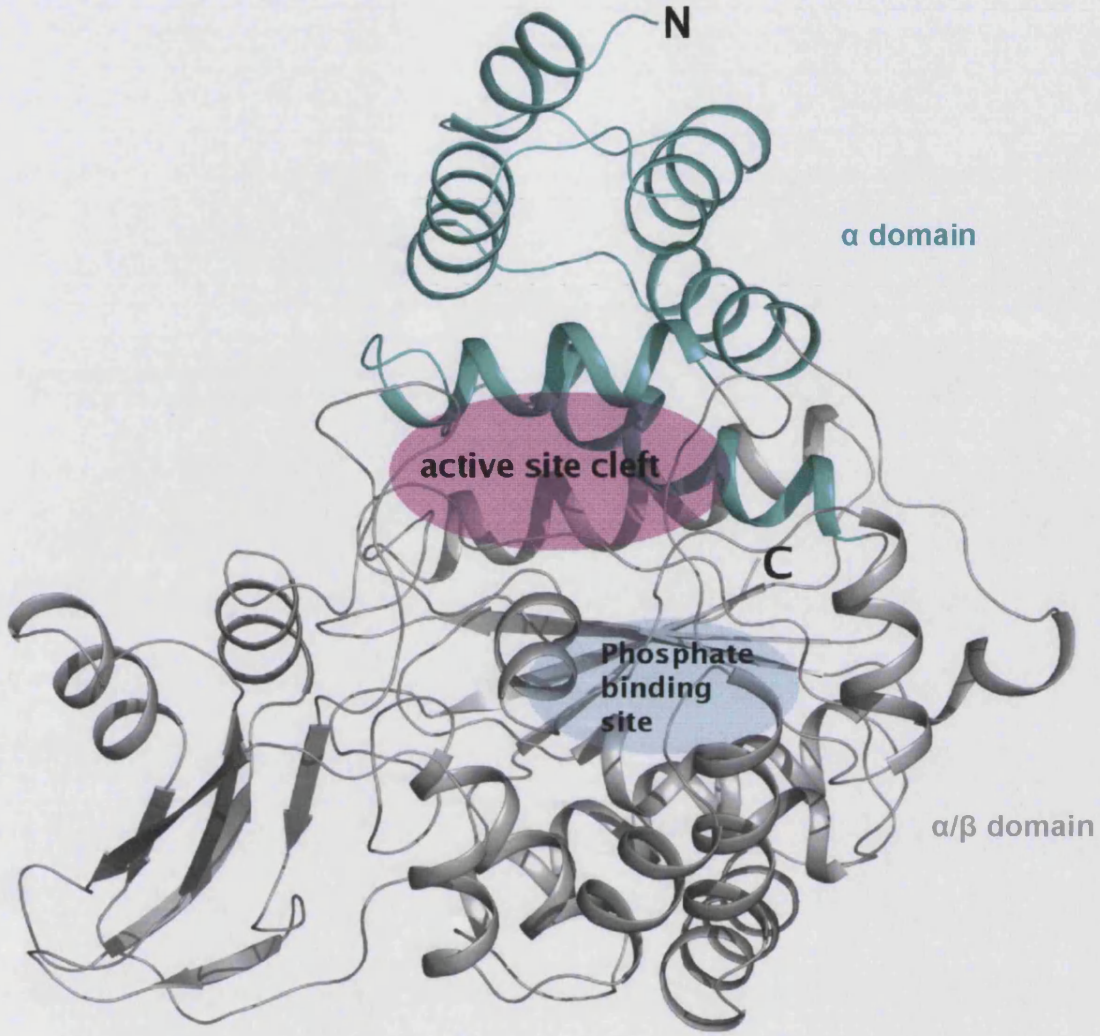


Figure 3.10: Cartoon representation of the secondary structure of hTP. The two domains are colour-coded: α domain and mixed α/β domain. The active site cleft, located at the interface between the two domains, is highlighted in pink whereas the phosphate binding site, located in the mixed α/β domain is highlighted in blue.

3.5.4 Domain movement in PyNPs

Early work on EcTP revealed that the distance between the substrate binding site and the phosphate binding site was too large (8Å) for the nucleophilic attack to occur (Pugmire *et al.*, 1998a). Hence a type of domain movement was proposed to take place upon binding

of substrates, where both α and α/β domains move as rigid bodies (Figure 3.11). The proposed sequence of events initiates with the binding of phosphate, which triggers the movement of the mixed α/β domain by 8°. This is followed by the binding of thymidine, which triggers a further domain movement, to a total of 21°, and leads to the fully closed (active) enzyme conformation (Pugmire and Ealick, 1998).

Three loop regions (L1, L2 and L3) 'bridge' the α and the mixed α/β domains. These three regions were proposed to confer flexibility for domain movement, which occurs upon occupation of the phosphate binding site and the active site. Another region of flexibility was proposed to be a seven residue loop (the hinge region) that extends across the active site cleft and appears to act as a 'latch' that forms a network of interactions with both domains and 'covers' the active site cleft (Pugmire and Ealick, 1998).

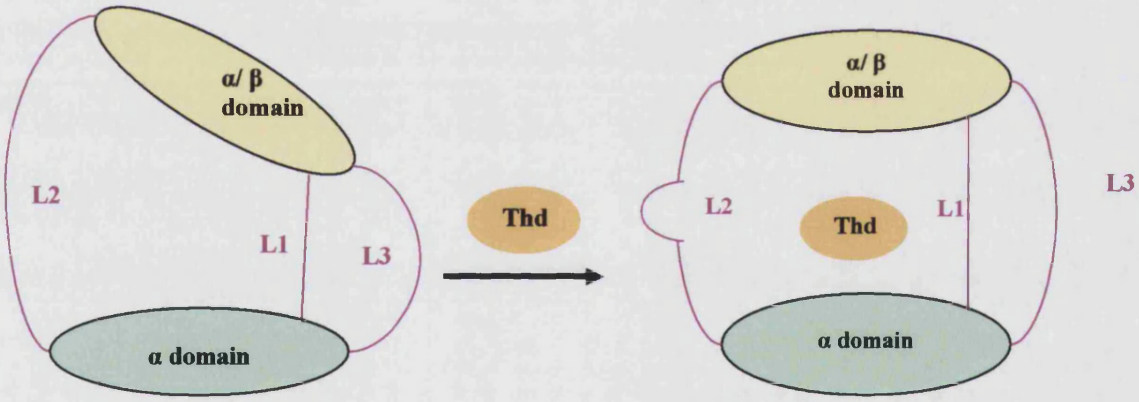


Figure 3.11: Schematic representation of the proposed movement of thymidine phosphorylase upon substrate binding. Thd represents thymidine, the natural substrate for thymidine phosphorylase. The three loops (L1, L2 and L3) are proposed to allow the rigid body movement that brings the two domains closer together for catalysis to occur. Shown in yellow is the mixed α/β domain and in light green is the α domain (Figure adapted from: Pugmire and Ealick, 1998).

Both structures of hTP complexes determined to date are in the closed conformation (1UOU, Norman *et al.*, 2004; 2J0F, El Omari *et al.*, 2006). The absence however of phosphate in both complexes of hTP might indicate that occupation of the active site alone, without prior occupation of the phosphate binding site, suffices to induce the 'closed' conformation of the enzyme (Norman *et al.*, 2004; El Omari *et al.*, 2006). All data currently available that support domain movement are based on structures of EcTP. Further structural data of an 'open' (unbound) form of hTP would provide valuable information on the two states of the enzyme and would indicate under which conditions domain movement is induced.

We have determined the structure of a complex of hTP with 5-iodouracil (5IUR), a small molecule inhibitor that greatly resembles the chemotherapeutic agent currently used, 5FU. Further structural information for hTP and possible conformational changes between an unbound (inactive) and a bound (active) conformation was obtained from the crystal structure of native hTP. Structural alignment of the bound and unbound structures of hTP provides new insights about potential domain movement and necessity of substrate/phosphate binding for inducing the closed conformation. To elucidate the role of certain active site residues for the enzymatic activity of hTP, mutagenesis experiments and enzymatic activity assays were carried out. New findings implicate residues which appear to have no direct interaction with the molecules bound in the active site of the enzyme and reveal which residues are indispensable for enzymatic activity.

A. Structural analysis

3.6 Materials and Methods

Crystals and data for hTP-5IUR complex and native hTP were obtained by Dr. Anastassios Papageorgiou. Crystals for hTP-5IUR and native hTP were grown with protein provided by Dr. Darell Sleep and Dr. Chris Finnis.

3.6.1 Crystallisation and X-ray data collection

Crystals for both native hTP and the hTP-5IUR complex were obtained by the hanging drop vapour diffusion method. Crystals of native hTP were grown in 0.1M sodium-acetate (pH 5.2), 25% PEG 4000 and 0.1M ammonium acetate and belonged to $P2_1$ space group. Crystals for the hTP-5IUR complex were obtained after overnight incubation of hTP with 100mM 5IUR at 4°C prior to equilibration over a reservoir solution containing 2.8-3.5 M sodium-formate (pH 8.7). The complex of hTP-5IUR crystallised in $P2_12_12_1$ space group. This condition is similar to that reported by Spraggon in the absence of any ligand (Spraggon *et al.*, 1993).

Data for hTP were collected in Elettra Synchrotron light source (λ=1 Å; Trieste, Italy). Data for hTP-5IUR complex were collected in SRS Daresbury (λ=1.48 Å; UK) and EMBL DESY (λ=0.84; Å Hamburg, Germany) and the collected data sets were merged together. Data collection was carried out at cryogenic temperatures.

3.6.2 Structure determination and refinement

Phases for the determination of the two hTP crystal structures were obtained by molecular replacement method using the program MOLREP (Vagin and Teplyakov, 1997). The coordinates of human thymidine phosphorylase (PDB code 1UOU; Norman *et al.*, 2004) were used as the search model for the hTP-5IUR complex. hTP-5IUR complex crystallised in $P2_12_12_1$ spacegroup with unit cell dimensions a=61.9, b=67.3 and c=212.3Å. Matthew's coefficient was calculated to be 2.3 $Å^3/D$ and the Wilson B-factor was calculated to be 46.7$Å^2$. The structure of the complex was initially refined using CNS (Brünger *et al.*, 1998). Refinement consisted of rigid body refinement (initial cycles of

refinement) as the model came directly from molecular replacement. The CNS task file *'rigid.inp'* was used to optimise the positions of the two monomers in the asymmetric unit. The result of rigid body refinement was a new coordinate file (rigid.pdb), which was used as input for simulated annealing that uses torsion angle dynamics to improve the model. The CNS task file *'anneal.inp'* was used to perform simulated annealing refinement. A starting temperature of 1000K was initially used since the model was identical to the target molecule. A slowcool protocol, applied during simulated annealing, heated up the molecule to overcome all energy barriers and then slowly cooled it down to approach energy minima. Examination of the maps revealed two loop regions, where the main chain electron density was broken: residues 237 to 238 and 406 to 415. To improve the phases, density modification was carried out. Density modification at medium resolutions, with multiple copies of the molecule in the asymmetric unit, has been shown to improve phases by reducing model bias. It calculates phase probability distribution for the model and the experimental amplitudes. The variations in electron density in the asymmetric unit are used to calculate a mask, which defines the protein/solvent boundary, and an electron density map. Density modification also provides an output reflection file of the observed amplitudes and reconstituted amplitudes. Three CNS task files were used for this purpose: *'model_phase.inp'*, *'density_modify.inp'* and *'fourier_map.inp'*. The electron density map calculated using the phases from density modification revealed electron density for the two loop regions. Once gross errors were corrected for, a round of rigid body refinement and simulated annealing was carried out to correct any errors introduced during manual model building and to optimise the fit of the retraced region. The R-factors improved, indicating correct changes were being made to the model. Towards the final steps of refinement of the hTP-5IUR complex structure torsion angle molecular dynamics coupled with temperature during simulated annealing were used as it allowed for sampling of several conformations at higher temperatures, avoiding the trapping of structures in false minima. For this purpose Refmac was used to incorporate TLS (Winn *et al.*, 2001) thermal refinement and each molecule within the asymmetric unit was defined as a TLS tensor group for the purpose of refinement. Refinement proceeded in 15 cycles with atomic residual isotropic B-factors set at 35Å^2, followed by 10 cycles of individual atomic restrained refinement. Once refinement of the structure

was completed, coordinates for 5-iodouracil were retrieved from the HIC-Up server. Water molecules were detected using the '*waterpick.inp*' input file from CNS (Brünger *et al.*, 1998) and were examined manually, to ensure no peaks were mistakenly picked as water molecules. Water-pick was carried out after the introduction of the inhibitor so that waters would not falsely occupy the inhibitor density. Water refinement included a total of 20 steps of minimisation and B-factor refinement. The minimum distance allowed between water molecules and any atom was set to 2.6Å, and the maximum distance was set to 4.0Å. A total of 89 waters were detected for the dimer of hTP-5IUR complex. The structure was refined to a crystallographic R-factor (R_{cryst}) of 23.9% and free R-factor (R_{free}) of 29.6% (Table 3.1).

Native hTP crystallised in $P2_1$ spacegroup with unit cell dimensions of a=101. 0, b=77.1, c=103.1Å and α=90°, β=98° and γ=90°. Matthew's coefficient was calculated to be 1.8 $Å^3/D$ using the CCP4 interface and Wilson B-factor was calculated to be 31.6 $Å^2$. For the structure of native hTP the search model used was the previously solved structure of hTP-5IUR complex. The structure was refined using CNS (as described for hTP-5IUR complex structure) to an R_{cryst} of 25.21% and an R_{free} of 31.6% (Table 3.1; Brünger *et al.*, 1998). Although the resolution was lower in this case no 'gaps' were identified in the main chain density. Due to the low observation-to-parameter ratio, individual B-factor refinement proved to be problematic. The use of anisotropic-fixed-isotropic option in the *'bindividual.inp'* file from CNS (Brünger *et al.*, 1998) allowed the correction of individual B-factors. Detection of water molecules was carried out using the '*waterpick.inp*' input file of CNS (Brünger *et al.*, 1998). Water refinement included a total of 20 steps of minimisation and B-factor refinement. The minimum distance allowed between water molecules and any atom was set to 2.6Å, and the maximum distance was set to 4.0Å. A total of 50 water molecules were detected for the native hTP structure. Programs employed for structure analysis included Molprobity (Lovell *et al.*, 2003), Procheck (Laskowski *et al.*, 1993), HBPLUS (McDonald and Thornton, 1994) and CONTACT (CCP4).

Table 3.1 : Statistics for data collection, processing and refinement

	hTP-5IUR complex	*hTP*
Spacegroup	$P2_12_12_1$	$P2_1$
Unit Cell Dimensions (Å)	a=61.9, b=67.3,c=212.3	a=101.0, b=77.1, c=103.1 β=98°
Resolution Range (Å)	50-2.5	38.35-3.0
Total reflections measured	371493	85968
Unique reflections measured	31655	31746
R_{merge} [a](%) (outermost shell)	15.60 (36.7)	14.9 (38.6)
I/σ(I)	4.5 (3.6)	4.9 (1.8)
Completeness (%)	97.6 (99.7)	89.9 (81.9)
R_{cryst} [b](%)	25.1	24.8
R_{free} [c](%)	31.0	31.4
R.m.s. deviation from ideality		
In bond lengths (Å)	0.007	0.007
In bond angles (deg.)	1.3	1.4
Average B-factor (Å^2)		
Overall	41.2	30.4

[a] $R_{symm} = \Sigma_h\Sigma_i[|I_{(h)} - \langle I_{(h)}\rangle|/\Sigma_h\Sigma_i I_{i(h)}]$, where $I_{i(h)}$ is the ith measurement and $\langle I_{(h)}\rangle$ is the weighted mean of all measurements of $I_{(h)}$

[b] $R_{cryst} = \Sigma_h||F_o| - |F_c||/\Sigma_h|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h.

[c] R_{free} is equal to R_{cryst} for a randomly selected 5% subset of reflections not used in refinement.

3.7 Results and discussion

3.7.1 Quality of native hTP and hTP-5IUR complex structures

The crystal structure of hTP-5IUR complex was determined at 2.5Å resolution with a crystallographic R-factor of 25.1% and an R_{free} of 31%. Two copies of the hTP-5IUR complex were identified in each asymmetric unit, with one molecule of 5IUR bound in the active site of each hTP monomer and no phosphate ions present in the phosphate binding sites (Figure 3.12).

Of the two copies present in the asymmetric unit, the second was slightly disordered with weak electron density in some areas. Arg342 in chain B has been modeled as alanine due to lack of visible electron density beyond C_β atom. The final structure contains a total of 89 water molecules and has 87.1% of the residues in the favourable regions of the Ramachandran plot (Figure 3.13; Ramachandran *et al.*, 1963). The two hTP monomers dimerise at the N-terminal domain, as previously reported by Norman and co-workers (Norman *et al.*, 2004).

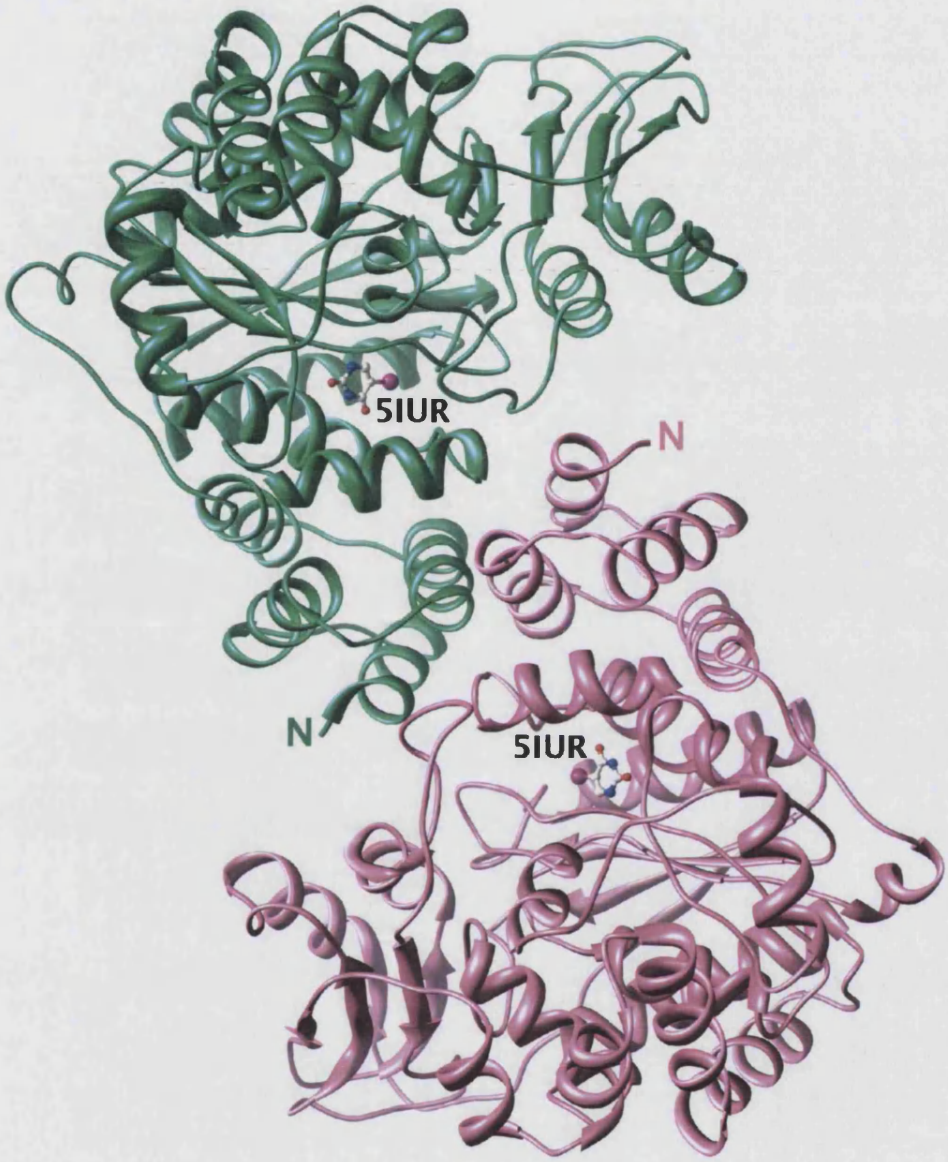


Figure 3.12: Cartoon representation for the dimer of hTP-5IUR. Each monomer is coloured differently: monomer A is coloured in pink and monomer B is coloured in green. One copy of 5IUR is bound in the active site cleft of each of the monomers. Dimerisation occurs at the N-terminal α-helical domain, as previously reported for other members of the PyNP superfamily.

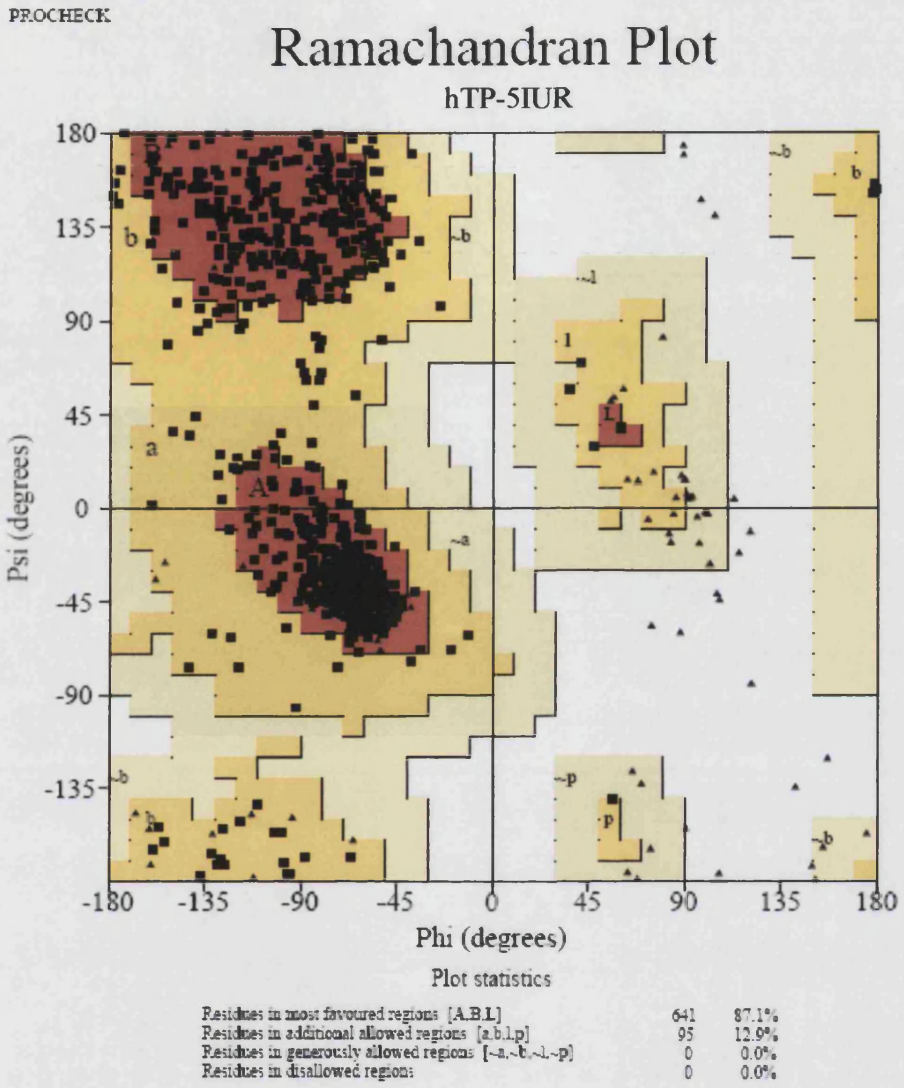


Figure 3.13: Ramachandran plot for the structure of hTP-5IUR complex. A total of 87.1% of the residues are on the most favourable regions, while 12.9% of the residues are in the additional allowed regions of the Ramachandran plot. No residues are in the generously allowed or disallowed regions.

The crystal structure of native hTP was determined at 3.0Å resolution with an R_{cryst} of 25.21% and an R_{free} of 31.6%. Four copies of hTP were identified in each asymmetric unit forming biological dimers with the symmetry related molecules (Figure 3.14).

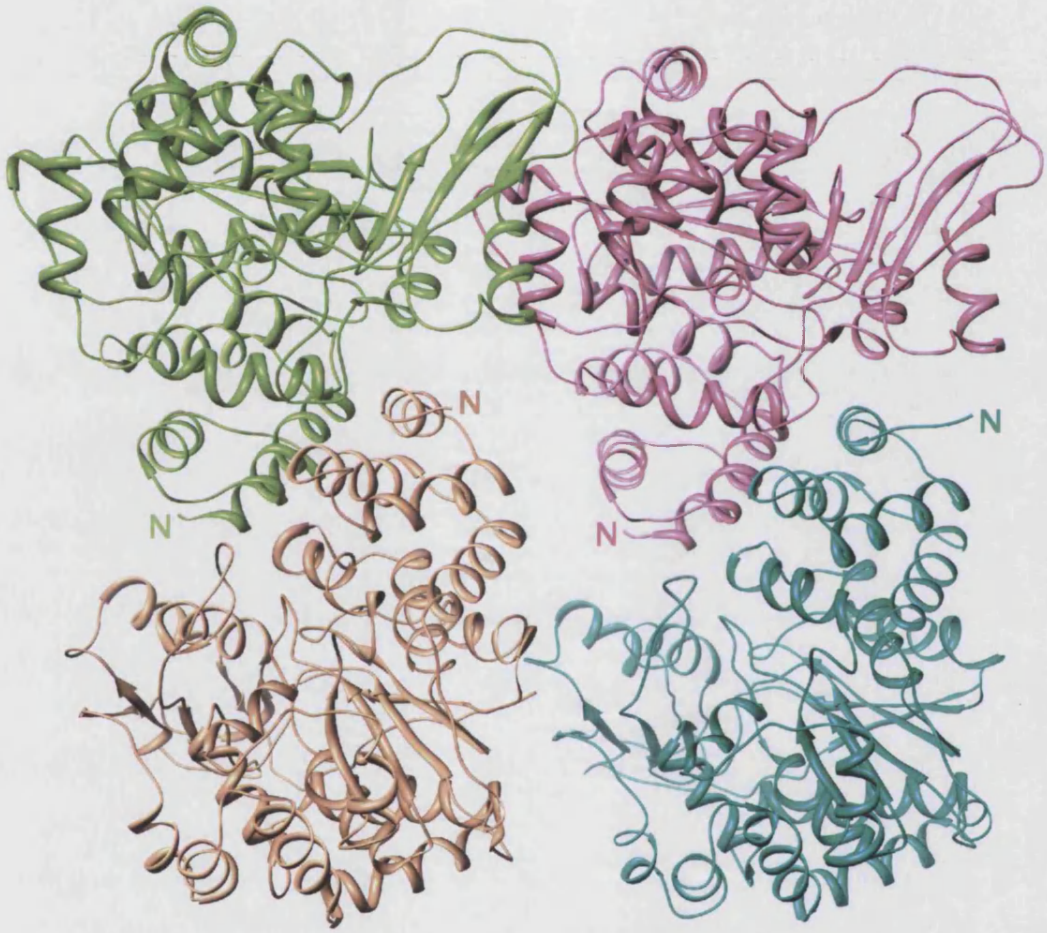


Figure 3.14: Cartoon representation for the native structure of hTP. Four molecules were present in each asymmetric unit, forming two dimers. The two dimers are illustrated here: monomer A is coloured in cyan and monomer B in pink. The second dimer is that of monomer C, coloured in golden, and monomer D, coloured in green.

For the three of the four monomers of native hTP in the asymmetric unit electron density was observed for residues 34-479, while electron density for residues 32 and 33 was observed for one of the monomers. In this crystal structure the active site of the enzyme is not occupied and neither is the phosphate binding site. The final structure has a total of 50 water molecules. Examination of the Ramachandran plot (Ramachandran *et al.*, 1963; Figure 3.15) shows 79.9% of the residues in the most favourable regions.

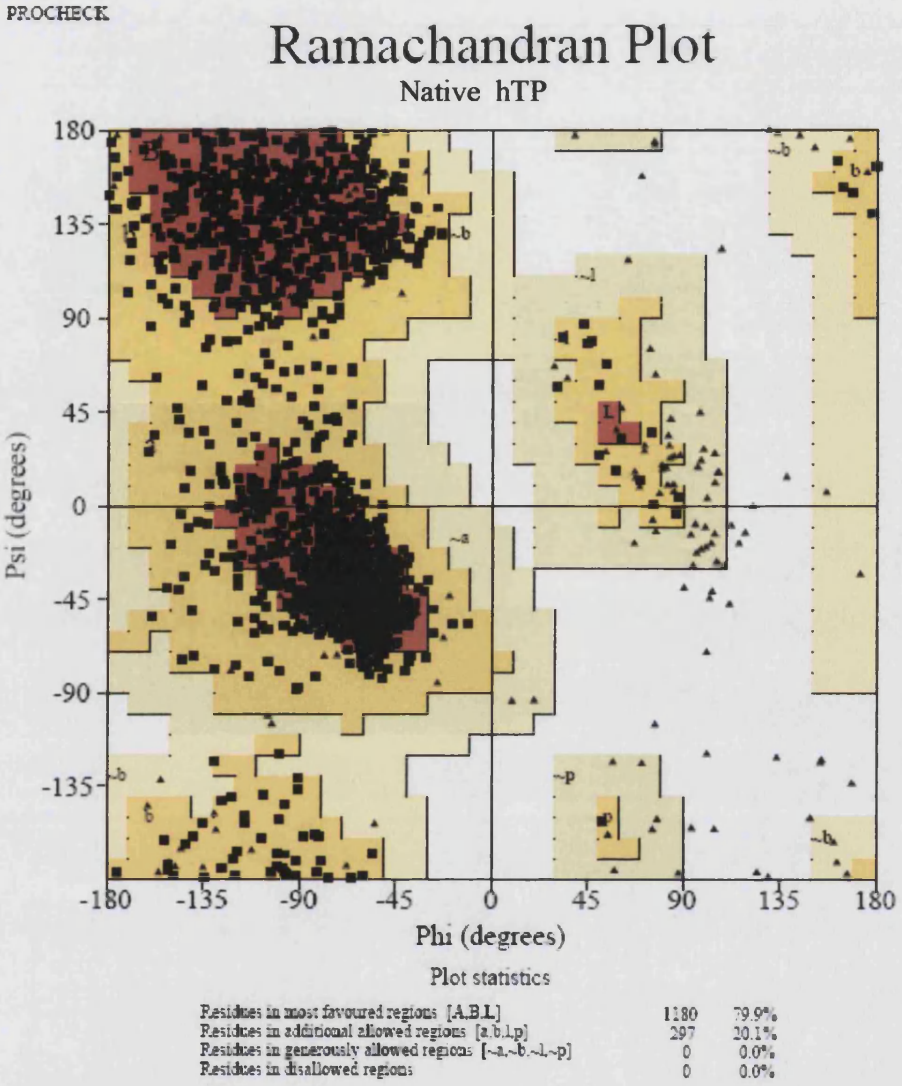


Figure 3.15: Ramachandran plot for the structure of native hTP. A total of 79.9% of the residues are on the most favourable regions, while 20.1% of the residues are in the additional allowed regions of the Ramachandran plot. No residues are in the generously allowed or disallowed regions.

The two different crystal forms observed for hTP-5IUR complex and native hTP appear to be the effect of crystal packing. The different crystallisation conditions under which hTP-5IUR and native hTP crystallised may have been the reason for the two different

crystal forms and space groups. Hence, hTP-5IUR complex crystallised in $P2_12_12_1$ spacegroup and native hTP crystallised in $P2_1$ spacegroup. In order to identify possible structural differences between the two forms a structural alignment of the dimer (hTP-5IUR) and the tetramer (native hTP) was carried out. More specifically, structural alignment of C^a atoms of hTP-5IUR (dimer) against each of the dimers of the native crystal structure (tetramer) yields an rmsd of ~1.3Å (SPDBV; Guex and Peitsch, 1997). The inter-molecular contacts observed in hTP dimer ($P2_12_12_1$) revealed some noticeable differences to those observed in the two dimers of the native hTP tetramer ($P2_1$). The total number of inter-molecular contacts identified for the dimer of hTP-5IUR is the same as those observed for one of the two dimers of native hTP, while the other dimer of native hTP tetramer has considerably less number of interactions at the interface. As no conformational changes can be detected for the two crystal structures, it is possible to assume that the differences observed are most likely due to crystal packing. The inter-subunit contacts observed in our two structures are similar to those reported previously for hTP structures hTP-TPI, (Norman *et al.*, 2004) and hTP-TDR (El Omari *et al.*, 2006).

3.7.2 Overall hTP structure

As the resolution of both native hTP and hTP-5IUR complex structures is low, it would not allow for a great amount of detail to be observed. At this resolution however it is possible to obtain information about the overall structure and observe conformational changes, if any. Hence each of the two structures as well as the comparison of the two provided valuable information.

No secondary structure variations were observed for native hTP and the hTP-5IUR complex structures from the previously reported structures of hTP (Norman *et al.*, 2004; El Omari *et al.*, 2006). Each monomer of hTP consists of a total of 18 α-helices and 13 β strands, organised in two domains: an α domain and a mixed α/β domain (Norman *et al.*, 2004). The α domain is composed of six helices (α1-α4 extending from residue 36 to 97 and α8-α9, extending from residues 193 to 223). The mixed α/β domain can be considered to have two sub-domains: the first is composed of a total of ten α-helices and six β-strands, organised in a central mixed β sheet (β1-β5, residues 115 to 272 and β13,

residues 473 to 477), which is surrounded by ten α-helices (α5-α7, residues 125 to 178 and α10-α16, residues 275 to 341). The second sub-domain of the mixed α/β domain consists of strands β6, β7, β9, β11 and β8, β10 and β12, forming two small β-sheets. Two α-helices (α17 and α18, residues 394 to 456) separate the two small β-sheets, near the C-terminal domain of hTP. Flexibility for the two domains of hTP has been proposed to be conferred by three polypeptide loops: L1, which connects α4 to β1, L2, which connects β3 to α8 and L3, which connects α9 to β4 (Figure 3.16; Norman *et al.*, 2004). A topology diagram illustrating the secondary structural elements of hTP is given below, Figure 3.16.

The N-terminal helices of the two monomers come together to form a coiled-coil interface and bury a total surface area of around 2010Å^2 upon dimerisation (Figure 3.12). Buried at the dimer interface are residues Gln35, Pro37, Ile40, Arg44, Asp45, Gln67, Gly68, Ala69, Gln70, Gly72, Ala73, Met76, Ala77, Arg79, Leu80, Ala201, Val204, Thr205, Ala206, Asp209, Glu403, Ser409 from both monomers, while residues Glu38, Arg41, Arg410, Ala411 and Gly412 are contributed only by monomer A and residues Ser65, Arg81, Ala406 and Arg408 from monomer B. Six potential hydrogen bonds and a hundred and sixty five van der Waals contacts are observed between the two monomers at the interface. Previous findings on EcTP (Walter *et al.*, 1990) suggested that the dimer interface was held mainly by hydrophobic interactions.

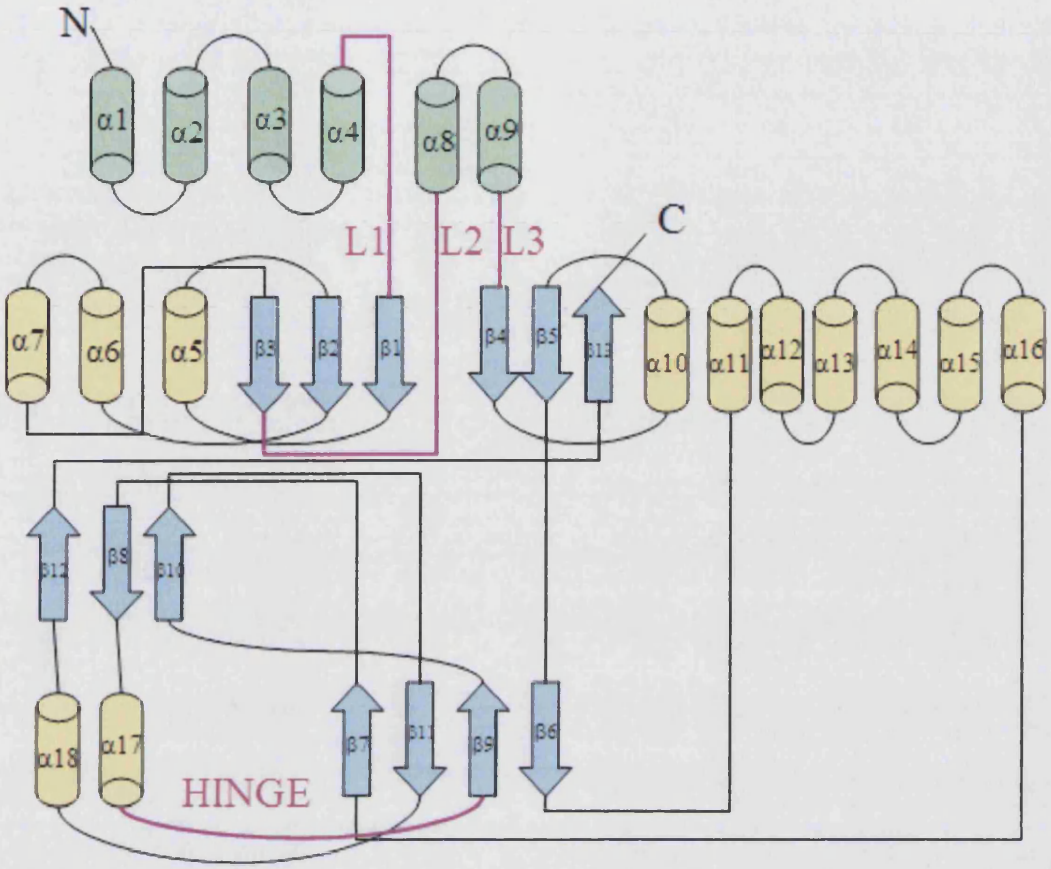


Figure 3.16: Topology diagram illustrating the secondary structure of hTP. The α helical is illustrated in light green, while helices which are part of the mixed α/β domain are illustrated in light yellow. The three polypeptide loops (L1, L2 and L3) and the hinge regions are highlighted in pink. The role of these flexible regions and their implication in the function of hTP is discussed in *3.7.4*. (Figure adapted from: Norman *et al.*, 2004).

3.7.3 Inhibitor binding and interactions of hTP with 5IUR.

The active site of hTP is located at the interface between the α and the mixed α/β domain, while the phosphate binding site is located below the active site cleft, in the mixed α/β domain (Walter *et al.*, 1990). Identification of putative hydrogen bonds was carried out with the program HBPLUS (McDonald and Thornton, 1994) while identification of

putative van der Waals interactions was carried out using the program CONTACT (CCP4). Complex formation results in a total of 747.4Å^2 surface area of each hTP monomer to be buried at the interface. The program *'buried_surface.inp'* of CNS (Brünger *et al.*, 1998) was used for the calculation of the buried surface. The residues that get buried upon complex formation were identified to be His116, Ser117, Thr118, Leu148, Tyr199, Ile214 and Ile218.

Previously proposed active site residues, conserved amongst species, are His116, Ser117, Leu148, Arg202, Val208, Ile214, Lys221 and Val241. Putative hydrogen bonds that stabilise the binding of 5IUR at the active site of hTP are detected for His116, Arg202, Ser217 and Lys221 in the hTP-5IUR structure (Figure 3.17). Almost 'opposite' to His116 in the active site cleft is Arg202, bonded *via* NH1 (3.0Å) and NH2 (2.8Å) to O4 of 5IUR. Ser217 OG appears to interact by means of hydrogen bonds to N3 (2.9Å) of 5IUR. Lys221 NZ forms a potential hydrogen bond with O2 of IUR (2.5Å). A close up view of the active site residues which form putative hydrogen bonds with 5IUR is shown in Figure 3.17. The potential hydrogen bonds as identified by HBPLUS (McDonald and Thornton, 1994) are presented in Table 3.2.

Of the active site residues, only Thr118 appears to form a putative hydrogen bond with a water molecule. No phosphate ion was present in the phosphate binding site of hTP-5IUR complex but a water molecule was detected at that particular position.

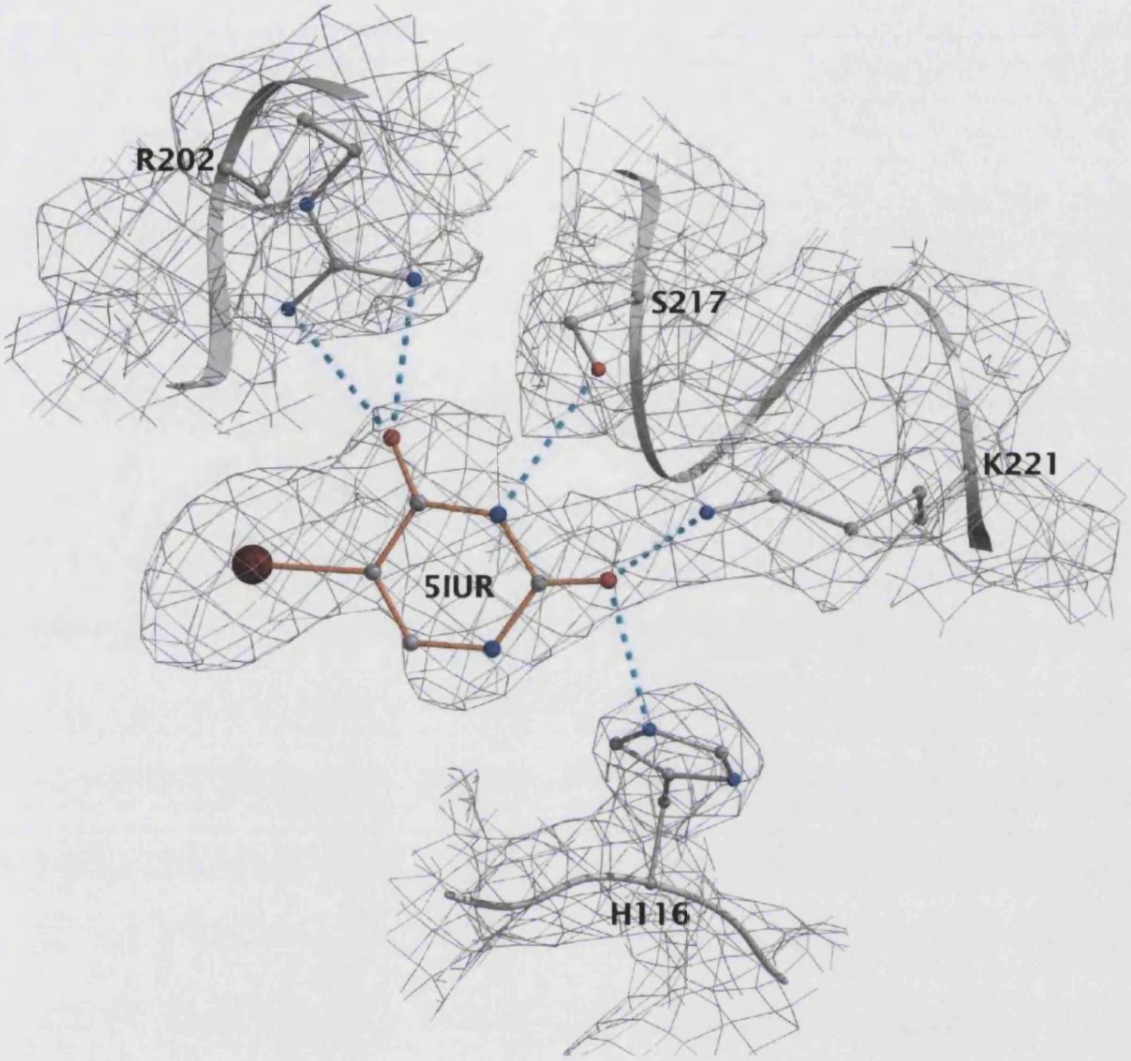


Figure 3.17: Close up of the active site residues forming putative hydrogen bonds with the bound 5IUR. Shown here are residues His116, Arg202, Ser217 and Lys221. Hydrogen bonds are illustrated in blue dashed lines. The map illustrated is a $2F_o$-F_c electron density map, contoured at 1σ level.

Table 3.2: Putative H-bonds for hTP-5IUR complex.

hTP Atom Chain_A	**B-factor (Å^2)**	**IUR 660 Y**	**B-factor (Å^2)**	**Distance D···A (Å)**	**Angle D-H···A (°)**
His116 NE2	38.44	IUR N1	45.56	3.0	-
His116 NE2	38.44	IUR O2	41.57	3.1	128.5
Arg202 NH1	26.38	IUR O4	43.80	3.0	142.9
Arg202 NH2	23.93	IUR O4	43.80	2.9	147.6
Ser217 OG	31.15	IUR N3	45.46	2.6	-
Lys221 NZ	42.00	IUR O2	41.57	2.5	152.2
hTP Atom Chain_B	**B-factor (Å^2)**	**IUR 661 Z**	**B-factor (Å^2)**	**Distance D···A (Å)**	**Angle D-H···A (°)**
Arg202 NH1	18.97	IUR O4	41.65	2.9	161.9
Ser217 OG	26.98	IUR N3	41.89	3.1	-

The upper limit for the donor-acceptor distance is 3.3Å. Contacts involving a sulphur atom have an upper limit of 3.6Å. The lower limit for the donor-acceptor angle is 120°. Bond angles are not given where the hydrogen position is ambiguous. The contact distances were calculated using the program HBPLUS (McDonald and Thornton, 1994).

Further stabilisation of 5IUR in the active site cleft of hTP is obtained by a network of electrostatic interactions. Residues identified to form this network of interactions include His116, Ser117, Thr118, Leu148, Tyr199, Arg202, Ile214, Ser217, Ile218, Lys221 and Val241. van der Waals interactions between hTP and 5IUR are presented in Table 3.3, as detected by CONTACT (CCP4, 1994).

Table 3.3: van der Waals contacts between hTP chains and 5IUR.

IUR 660 Y Atoms	hTP Chain_A residues	No. of contacts
IUR 660 Y N1	His116	1
IUR 660 Y C2	His116,Tyr199,Ser217,Ile218,Lys221	5
IUR 660 Y O2	His116,Lys221	2
IUR 660 Y N3	Ile218	1
IUR 660 Y C4	Leu148,Ile214,Ser217,Ile218	4
IUR 660 Y O4	Arg202,Ile214$^{(8)}$	9
IUR 660 Y C5	Leu148,Ile214	2
IUR 660 Y I	Leu148,Thr118$^{(2)}$,Ile214$^{(2)}$,Val241	6
IUR 660 Y C6	Ser117,Thr118$^{(2)}$,Leu148	4

IUR 661 Z Atoms	hTP Chain_B residues	No. of contacts
IUR661 Z N1	His116	1
IUR661 Z C2	His116,Tyr199,Ile218,Lys221	4
IUR661 Z O2	His116,Tyr199,Lys221	3
IUR661 Z N3	Ser217,Ile218	2
IUR661 Z C4	Leu148,Ile214,Ser217,Ile218	4
IUR661 Z O4	Arg202$^{(3)}$,Ile214$^{(8)}$	11
IUR661 Z C5	Leu148,Ile214	2
IUR661 Z I	Leu148,Thr118$^{(2)}$,Ile214$^{(2)}$,Val241	6
IUR661 Z C6	Leu148,Ser117,Thr118$^{(2)}$	4

The maximum allowed values of intermolecular distances are **C-C**, 4.1Å;**C-N**, 3.8 Å;**C-O**, 3.7 Å;**O-O**, 3.3 Å;**O-N**, 3.4 Å; **N-N**, 3.4 Å;**C-S**, 4.1 Å;**O-S**, 3.7 Å;**N-S**, 3.8 Å;**I-C**, 4.3 Å;**I-N**, 3.9 Å;**I-O**, 3.7 Å. Superscript numbers in parentheses represent the number of contacts made by the indicated residue. Contact distances were calculated using the program CONTACT (CCP4, 1994).

Conservation of the active site residues is supported by previous work on EcTP where residues His85, Arg171, Ser186 and Lys190 were identified to interact with the substrate

bound at the active site (Walter *et al.*, 1990; Pugmire *et al.*, 1998a). The role of one of these residues, His85, has attracted a lot of interest, as it was proposed to be indispensable for the maintenance of the enzymatic activity of TP (Mendieta *et al.*, 2004). Similarly the structure of BsPyNP indicated that residues interacting with the pseudouridine moiety bound at the active site, include Lys187, Ser183, Arg168, Tyr165 and His82 (Pugmire and Ealick, 1998). It is evident that these charged residues are forming a network of interactions, similar to that observed for the active site of hTP. The conservation of these residues amongst species is thus highlighting their significance in the catalytic activity of TP.

The binding of thymidine to the active site cleft of TP was described in detail by Walter and co-workers (Walter *et al.,* 1990) based on structures including 5-iodouracil, thymine and thymidine (available on the Protein Data Bank is the structure of EcTP in complex with sulphate and thymine; PDB ID: 1TPT). A surface representation of the active site cavity of EcTP (Walter *et al.,* 1990) where thymidine and phosphate were modeled, indicated a tyrosine residue (Y168 EcTP; Y199 hTP; Y165 BsPyNP) involved in interactions with the ribose moiety of thymidine. Similar interactions have not been described for the complexes of BsPyNP (Pugmire and Ealick, 1998) and hTP (Norman *et al.,* 2004; El Omari *et al.,* 2006). The lack of this particular interaction is most likely due to the size of the molecules bound in the active site of the BsPyNP and hTP complex structures; the smaller size of pseudouridine in BsPyNP (Pugmire and Ealick, 1998) and TPI and thymine in hTP (Norman *et al.,* 2004; El Omari *et al.,* 2006) probably does not allow for the total number of interactions to be observed.

Pugmire and Ealick first described the presence of a hydrophobic pocket within the active site cleft of the BsPyNP structure (Pugmire and Ealick, 1998). The residues described to compose this pocket in BsPyNP are Phe207, Val174, Ile180 and Leu114 where they appeared to surround the 5'-position of the pyrimidine ring, bound in the active site of BsPyNP (Pugmire and Ealick, 1998). Norman and co-workers (Norman *et al.*, 2004) later reported that the same hydrophobic pocket was identified in the structure of hTP-TPI complex. More specifically, residues Val208 and Ile214 of the α-domain and Leu148 and

Val241 of the mixed α/β-domain were proposed to stabilise the chlorine atom of TPI in a manner similar to that previously seen for BsPyNP and pseudouridine. In the structure of hTP in complex with 5IUR the same hydrophobic pocket appears to 'hold' the iodine atom of 5IUR, in a manner similar to that previously proposed. A close up view of the active site, highlighting the hydrophobic pocket is presented below; Figure 3.18.

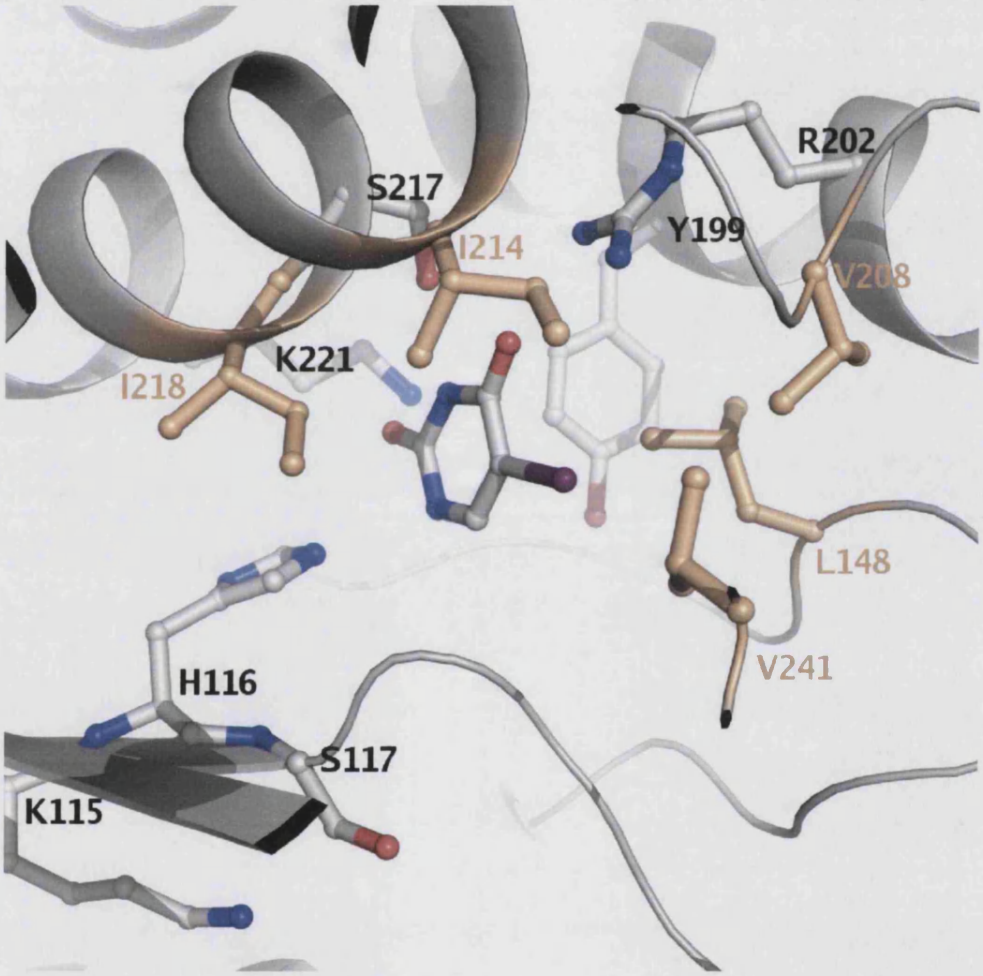


Figure 3.18: Close up view of the active site cleft, rotated by 90°. Shown here are the active site residues previously described to interact *via* hydrogen bonding interactions with 5IUR. Highlighted in golden are the residues that form the hydrophobic pocket, residues Leu148, Val208, Ile214, Ile218 and Val241.

Of the two α domain residues that are part of the hydrophobic pocket, Val208 is at a distance of 4.1Å and Ile214 is at a distance of 3.7Å from the iodine atom of 5IUR. Of the two α/β domain residues that form the hydrophobic pocket, Leu148 is at a distance of 3.9Å and Val241 is at a distance of 4.2Å from the iodine atom of 5IUR. His116, part of the active site cleft, interacts *via* its CE1 with O2 of 5IUR at a distance of 2.9Å. The carbonyl of Thr118 is located at a distance of 3.6Å from the iodine atom of 5IUR. It is evident that these residues, with the exception of Val241, which corresponds to Phe207 in BsPyNP, are conserved amongst species. Hence apart from the hydrogen bond network which is essential for stabilisation, the hydrophobic pocket also appears to play a crucial role.

3.7.4 Domain movement and comparison of bound and unbound hTP structures

Of the two substrates that are required for catalysis by hTP, phosphate is considered to bind first, followed by the binding of the substrate, thymidine. Domain movement (once the active site is occupied) is considered to be essential in bringing the substrates into close proximity for the reaction to occur. Hence a theory of domain movement has been proposed for the members of the PyNP superfamily.

Structural studies on EcTP revealed that the distance between the substrate binding site and the phosphate binding site appears to be too large (8Å) for reaction to occur (Pugmire *et al.*, 1998a). Modeling of thymidine in the substrate binding site of EcTP indicated that upon substrate binding the cavity between the two domains is occupied and no accessibility is permitted to the phosphate binding site (Walter *et al.*, 1990). Hence the two domains are proposed to come closer and reduce the distance between the two reactants. This movement is considered to shield the substrates from solvent access and trap the intermediates so that they are not released before completion of the reaction (Pugmire *et al.*, 1998a). Reducing the solvent access has been proposed to be of key importance for the reaction to occur: if the aqueous environment during phosphorolysis is in vast excess over phosphate, hydrolysis may compete with phosphorolysis (Pugmire *et al.*, 1998a).

When the active site of the enzyme is not occupied, it is proposed to acquire an 'open', inactive conformation. Upon occupation of both active and phosphate binding sites, domain movement is considered to occur as a rigid body movement for both α and α/β domains. A total of 21° rotation was observed for the α/β domain with respect to the α domain in BsPyNP when compared to an 'open' conformation of EcTP (Pugmire and Ealick, 1998). Adding to this flexibility are three polypeptide loops, L1, L2 and L3 which are proposed to assist domain movement. Of the three loops, L1 and L2 appear to undergo significant conformational changes when the 'open' and the 'closed' form of the enzyme were compared (Mendieta *et al.*, 2004). Additional information by Pugmire and Ealick indicates that of the three loops, L2 plays the most important role in bridging the gap between the two domains, while L1 and L3 are proposed to act as passive hinges between the two domains when movement occurs (Pugmire and Ealick, 1998). A key hydrogen bond, formed between the side-chain of His119 and the carbonyl oxygen of Gly208 has been proposed to form in the 'closed' conformation of the enzyme, when the two domains are in close proximity (Walter *et al.*, 1990; Pugmire and Ealick, 1998). Another area of flexibility was proposed to be the hinge region (residues 364 to 371 in BsPyNP, corresponding to residues 405 to 415 in hTP) that extends across the active site cleft and appears to act as a 'latch' (Pugmire and Ealick, 1998).

The structure of EcTP lead to the identification of a glycine-rich loop, preceded by a β-strand (β2; Figure 3.16) and followed by an α-helix (α6; Figure 3.16), which was proposed to be involved in the binding of the phosphate ion. This feature has been identified in other nucleotide-binding proteins (Dreusicke and Schulz, 1986). Due to its high content in glycine residues it is considered to be very flexible and aid the movement of the two domains (Pugmire *et al.*, 1998a). This glycine-rich loop region is conserved in all known TP structures, signifying its potential functional implication.

The two crystal structures of hTP reported to date, hTP-TPI (Norman *et al.,* 2004) and hTP-thymine (El Omari *et al.,* 2006) were both in the 'closed' (active) conformation (Norman *et al.*, 2004; El Omari *et al.*, 2006). Based on previous work (El Omari *et al.*, 2006) the findings of which suggested that the 'closed' form of the enzyme can be

induced solely by occupation of the active site, it is possible to argue that binding of 5IUR suffices to induce domain movement and lead to the 'closed' and hence active conformation of hTP.

More information about the proposed domain movement and its occurence in hTP was obtained by the native (unbound) structure of the enzyme. The crystal structure of unbound hTP provided some very controversial *albeit* interesting findings. Alignment of the C_α_chain of the two structures (native hTP and hTP-5IUR complex) indicates that they overall align well, with an rmsd of 0.7Å over 446 C_α atoms (Figure 3.19). The good alignment of the two structures and the low rmsd are indicative of the lack of conformational differences. This is opposed to all previously proposed theories according to which one would expect the unbound state of the enzyme to acquire an 'open' conformation.

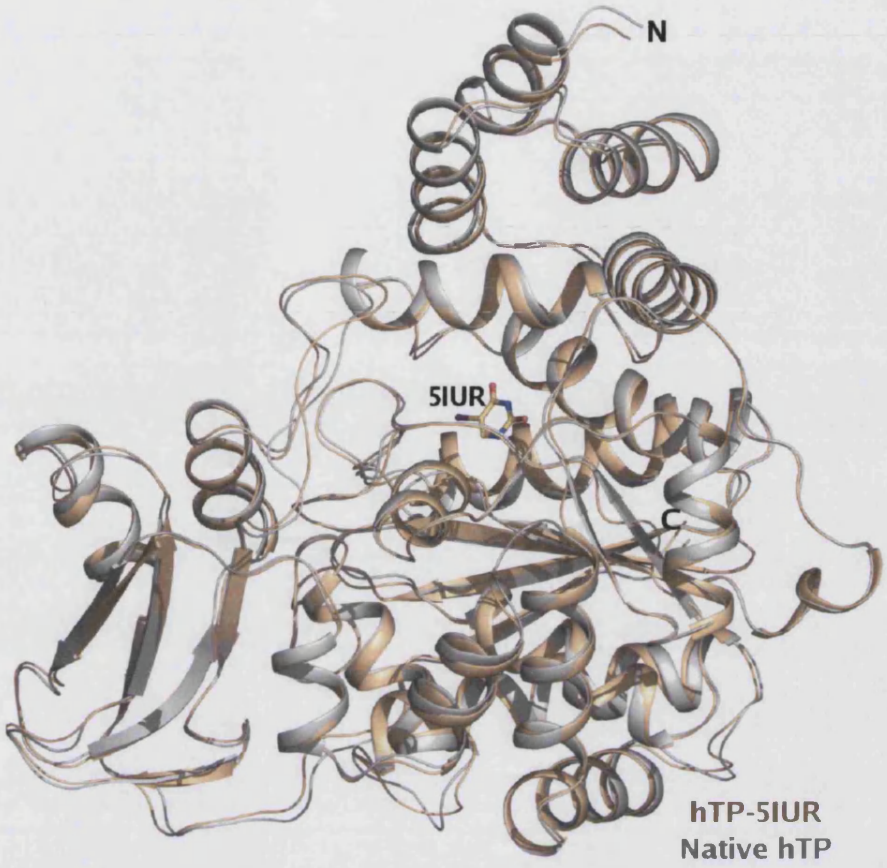


Figure 3.19: Structural alignment of native hTP (in grey) and hTP in complex with 5IUR (in golden). The two structures appear to align well with no areas of significant differences. The secondary structure features superimpose well for the two structures indicating that there is no domain movement that would separate a bound (open) from an unbound (closed) conformation, as expected according to the proposed theory of domain movement.

Adding to this observation were the results from the examination of the previously proposed areas of flexibility, loops L1, L2 and L3 as well as the hinge region (Figure 3.20). Structural alignment of the bound and unbound hTP structures revealed that all flexible loops align well in the two structures, again indicating that no conformational change, in these regions, exists between the bound and the unbound states of the enzyme.

The hinge region, also proposed to have a similar role, appears to have no major differences between the two structures; a small area at the beggining of this loop only appears to differ between the bound and unbound states of hTP (Figure 3.20). It is important however to note that this region of the loop is not located near the active site and hence such small difference would not be enough to explain the flexibility of the whole loop. The observed difference in the hinge region is most likely due to crystal packing. A structural alignment of bound and unbound hTP, highlighting the three polypeptide loops and the hinge region is presented in Figure 3.20.

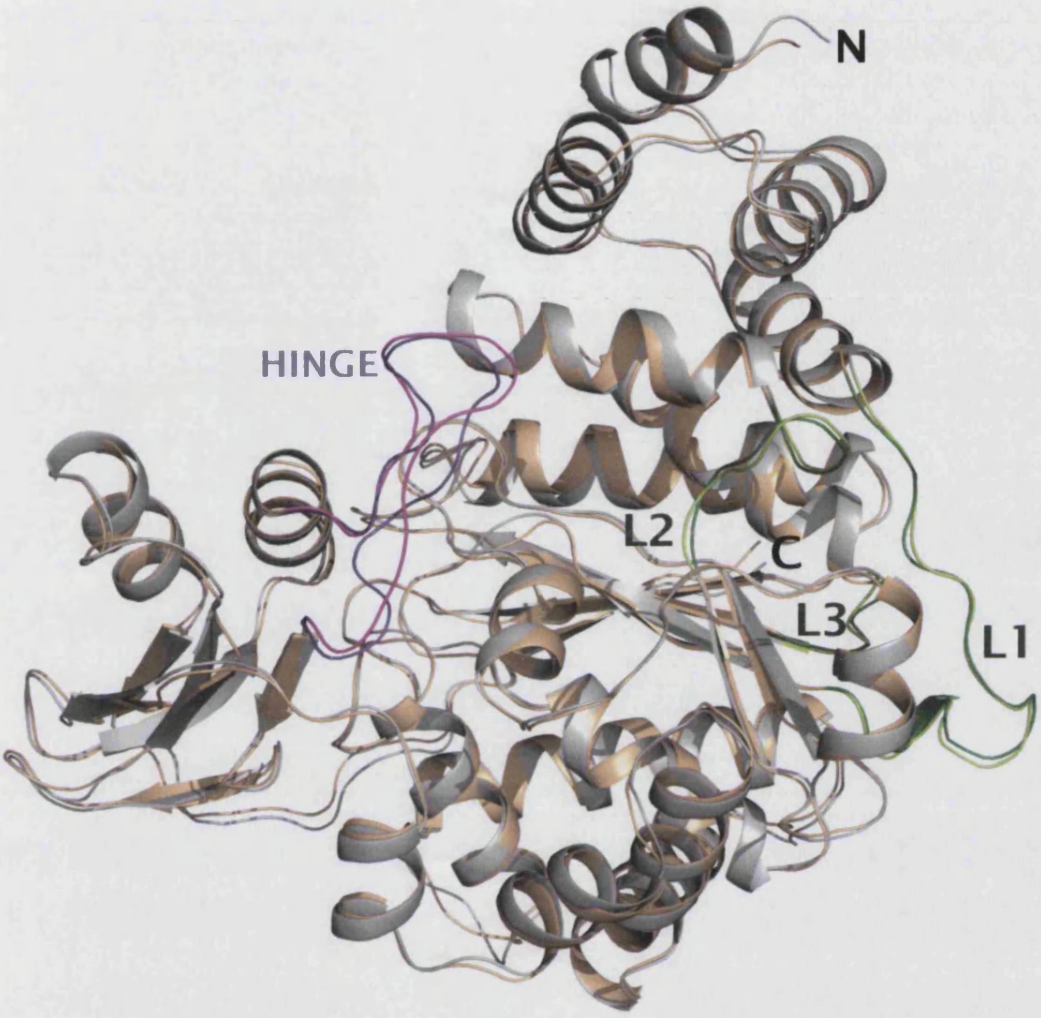


Figure 3.20: Structural alignment of native hTP (in grey) and hTP in complex with 5IUR (in golden), highlighting the three flexible polypetide loops L1, L2 and L3 (in green) and the hinge region (in magenta). It is evident that even in these areas of flexibility the two structures, bound and unbound hTP, align well.

Previous work on EcTP and BsPyNP lead to the identification of a key feature of the 'closed' enzyme conformation. A key hydrogen bond, formed between the side-chain of His119 and the carbonyl oxygen of Gly208 (in BsPyNP) has been proposed to form when phosphate binds to the phosphate binding site, leading to the 'closed' conformation of the enzyme (Walter *et al.*, 1990; Pugmire and Ealick, 1998). The residues involved in this hydrogen bond, correspond to His150 and Ala239 in hTP. Interestingly, an identical hydrogen bond was identified in both bound and unbound structures of hTP (Figure 3.21), possibly indicating that native hTP is also present in the closed conformation, eventhough no substrate or phosphate are present in the active site cleft. It is also possible that this hydrogen bond is not a key feature of only the closed (bound-active) state of the enzyme.

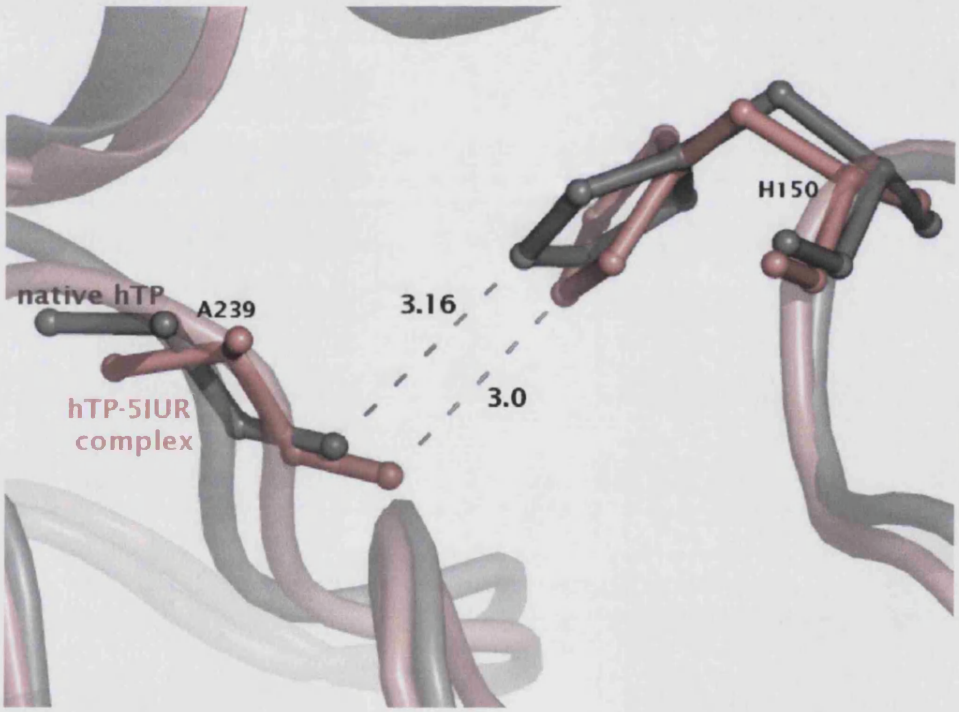


Figure 3.21: Shown in this figure is the key hydrogen bond proposed to form in the bound (active) conformation of TP. This hydrogen bond was identified in the structures of both bound hTP (shown in pink) and in native hTP (shown in grey).

Another common feature of the two structures of hTP is the presence of a water molecule in the phosphate binding site (Figure 3.22). It is unlikely that solely the binding of water molecules would trigger the enzyme to acquire its half-closed conformation (proposed 8° rotation upon phosphate binding). It is likely that the proposed 8° rotation would be observed if phosphate (and not water) would bind at the phosphate binding site, as hTP is a potent phosphorylase and not a hydrolase. Hence, binding of phosphate could 'activate' the phosphorolytic activity of hTP and initiate the phosphorolytic reaction.

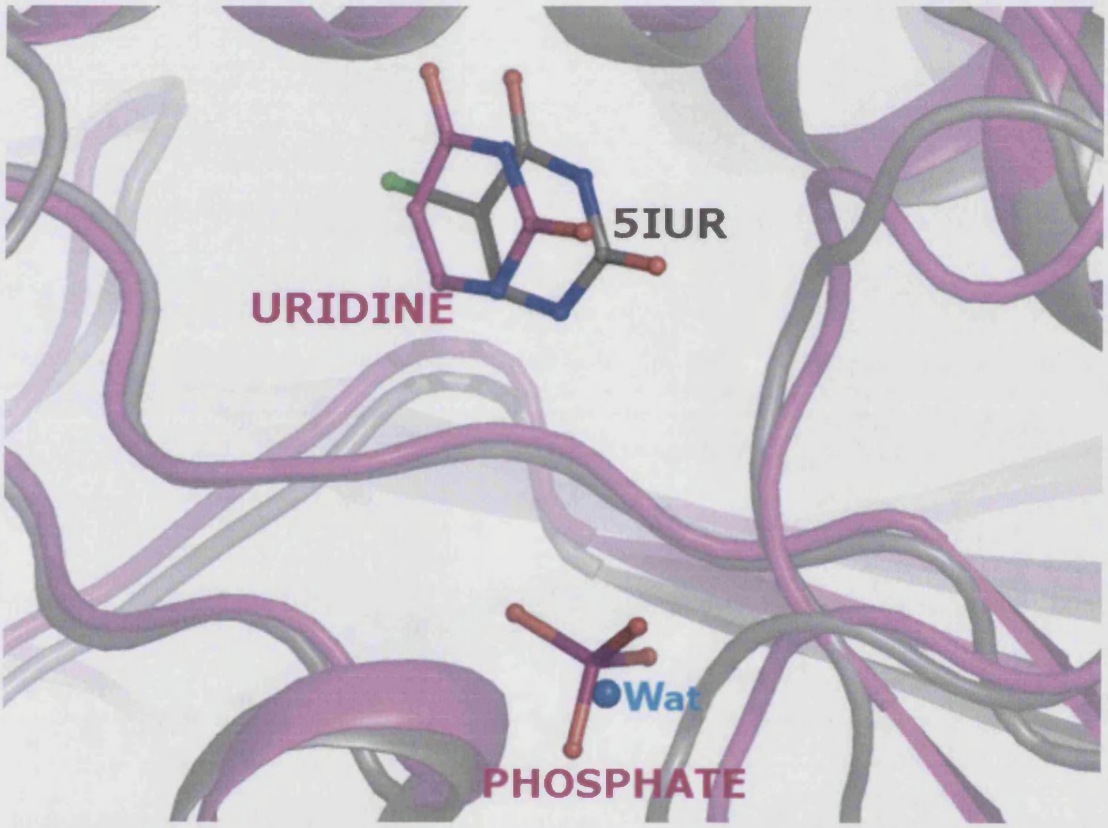


Figure 3.22: Shown here is a structural alignment (full length) of hTP-5IUR (in grey) against BsPyNP (in magenta). In the structure of BsPyNP, a phosphate ion is present in the phosphate binding site. In the structure of hTP-5IUR a water molecule occupies the phosphate binding site.

Rick and co-workers (Rick *et al.*, 1999) performed quantum mechanical calculations using the structure of EcTP in order to elucidate the mechanism of the catalytic activity of the enzyme. A mechanism of catalysis, not based on the crystal geometry, suggests that the reaction occurs in two steps: once both phosphate and substrate are bound, there is a direct nucleophilic attack of the phosphate on the glycosidic bond, resulting in its cleavage. This is then followed by phosphorylation of the ribose ring, where the reaction is completed. There is limited information about product release although another conformational step would probably be required before that occurs (Rick *et al.*, 1999).

Comparison of the two states of hTP (bound and unbound) revealed that presence of substrate or substrate analogues is not required for the enzyme to acquire the 'closed' conformation. Overall comparison reveals no conformational differences. It is quite likely that domain movement does not occur in hTP. It should be mentioned that hTP is larger to EcTP by 42 residues and to BsPyNP by 49 residues, a small difference which would not account for such a significant conformational change, as proposed to exist between a bound and an unbound state. It is possible however that the closed corformation observed for unbound hTP, is favoured by crystal packing, while different forms of the enzyme exist in solution. As no structure of transition states of the enzyme have been reported to date, it is not possible to state with confidence whether domain movement does or does not occur in hTP.

3.7.5 Comparison of hTP with other known TP structures.

The homology between hTP, EcTP and BsPyNP was calculated to be ~41% (Walter *et al.,* 1990; Brodsky *et al.*, 1992). Most of the α domain and part of the mixed α/β domain are conserved between the three PyNPs. High conservation can be observed especially for the active site and phosphate binding sites (Figure 3.23).

(A)

```
BsPyNP          ---------------------------------MRMVDLIAKKRDGKALTKEEIEWIVRG
hTP             MAALMTPGTGAPPAPGDFSGEGSQGLPDPSPEPKQLPELIRMKRDGGRLSEADIRGFVAA
EcTP            --------------------------------MFLAQEIIRKKRDGHALSDEEIRFFING
                                                                    ...
BsPyNP          YTNGDIPDYQMSALAMAIYFRGMTEEETAALTMAMVQSGEMLDLSS--IRGVKVDKHSTG
hTP             VVNGSAQGAQIGAMLMAIRLRGMDLEETSVLTQALAQSGQQLEWPEA-WRQQLVDKHSTG
EcTP            IRDNTISEGQIAALAMTIFFHDMTMPERVSLTMAMRDSGTVLDWKSLHLNGPIVDKHSTG
                                                .
BsPyNP          GVGDTTTLVLGPLVASVGVPVAKMSGRGIGHTGGTIDKLESVPGFHVEISKDEFIRLVNE
hTP             GVGDKVSLVLAPALAACGCKVPMISGRGLGHTGGTLDKLESIPGFNVIQSPEQMQVLLDQ
EcTP            GVGDVTSLMLGPMVAACGGYIPMISGRGLGHTGGTLDKLESIPGFDIFPDDNRFREIIKD
                                .  .     .     .  ..  .
BsPyNP          NGIAIIGQTGDLTPADKKLYALRDVTATVNSIPLIASSIMSKKIAAGADAIVLDVKTGAG
hTP             AGCCIVGQSEQLVPADGILYAARDVTATVDSLPLITASILSKKLVEGLSALVVDVKFGGA
EcTP            VGVAIIGQTSSLAPADKRFYATRDITATVDSIPLITASILAKKLAEGLDALVMDVKVGSG
                 .
BsPyNP          AFMKKLDEARRLARVMVDIGKRVGRRTMAVISDMSQPLGYAVGNALEVKEAIETLKGNGP
hTP             AVFPNQEQARELAKTLVGVGASLGLRVAAALTAMDKPLGRCVGHALEVEEALLCMDGAGP
EcTP            AFMPTYELSEALAEAIVGVANGAGVRTTALLTDMNQVLASSAGNAVEVREAVQFLTGEYR

BsPyNP          -HDLTELCLTLGSHMVYLAEKAPSLDEARRLLEEAIRSGAAIAAFKTFLAAQGGDASVVD
hTP             -PDLRDLVTTLGGALLWLSGHAGTQAQGAARVAAALDDGSALGRFERMLAAQGVDPGLAR
EcTP            NPRLFDVTMALCVEMLISGKLAKDDAEARAKLQAVLDNGKAAEVFGRMVAAQKGPTDFVE

BsPyNP          DL---------DKLPKAAYTSTVTAAADGYVAEMAADDIGTAAMWLGAGRAKKEDVIDLA
hTP             ALCSGSPAERRQLLPRAREQEELLAPADGTVELVRALPLALVLHELGAGRSRAGEPLRLG
EcTP            NYA--------KYLPTAMLTKAVYADTEGFVSEMDTRALGMAVVAMGGGRRQASDTIDYS

BsPyNP          VGIVLHKKIGDRVQKGEALATIHSNR-PDVLDVKEKIEAAIRLSPQP-VARPPLIYETI-
hTP             VGAELLVDVGQRLRRGTPWLRVHRDGPALSGPQSRALQEALVLSDRAPFAAPSPFAELV-
EcTP            VGFTDMARLGDQVDGQRPLAVIHAKDENNWQEAAKAVKAAIKLADKA-PESTPTVYRRIS

BsPyNP          V----
hTP             LPPQQ
EcTP            E----
```

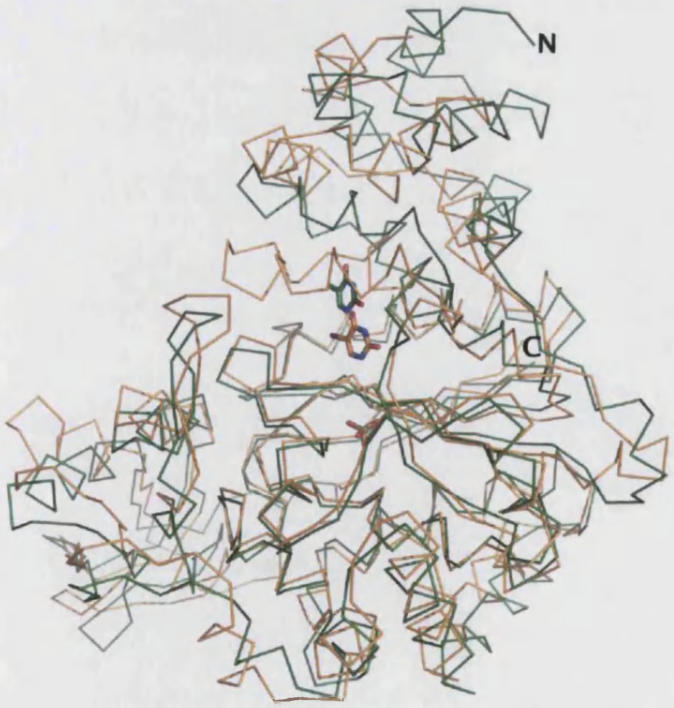
(B)

Species	Conserved residues amongst species (in the order highlighted in the sequence alignment)											
BsPyNP	K81	H82	S83	L114	Y165	R168	V174	I180	S183	I184	K187	F207
hTP	K115	H116	S117	L148	Y199	R202	V208	I214	S217	I218	K221	V241
EcTP	K84	H85	S86	L117	Y168	R171	V177	I183	S186	I187	K190	F210

Figure 3.23: **(A)** Full length sequence alignment of hTP, EcTP and BsPyNP. Conserved residues amongst the three PyNPs marked (•) and highlighted in pink. Shown in cyan is one of the members of the hydrophobic pocket, Val241 in hTP which corresponds to F207 in BsPyNP and F210 in EcTP. **(B)** The table illustrates the conserved residues, in the order highlighted in the sequence alignment, with the respective numbering for each of the three PyNPs examined; hTP, EcTP and BsPyNP.

Structural comparison of hTP, EcTP and BsPyNP gave an insight into the structural similarity of the three PyNPs. The overall structural pattern, typical for members of the PyNP superfamily is observed in the three enzymes. Structural alignments for all structures were calculated for the backbone C^{α} atoms, using the program SPDBV (Guex and Peitsch, 1997). Alignment of native hTP against EcTP-thymine complex (PDB ID: 1TPT) yielded an rmsd of 1.5Å (Figure 3.24), while alignment of hTP-5IUR complex against EcTP-thymine complex yielded an rmsd of 1.6Å. When the unbound structure of EcTP (PDB ID: 1AZY) was aligned against native hTP the rmsd obtained was 1.7Å, equal to that obtained for its alignment against hTP-5IUR complex (Table 3.4; Figure 3.24 to 3.26).

Figure 3.24: Structural alignment of native hTP (in grey) against EcTP-thymine complex (in green). Differences between the aligned EcTP structures and hTP structures are mainly identified for the α helical domain, while the mixed α/β domain appears to have fewer differences amongst the aligned structures. Bound in the active site cleft of EcTP is thymine and in the phosphate binding site is a sulphate ion.



Structural alignment of hTP-5IUR complex (in orange) against bound EcTP (in green). Again differences are mainly observed in the α helical domain.

Bound in the active site cleft of EcTP is thymine and in the phosphate binding site is a sulphate ion. Superposition of bound EcTP and hTP-5IUR structures reveals that the molecules bound in the active site cleft do not align perfectly. This could possibly be due to the α helical domains not aligning well for the two structures.

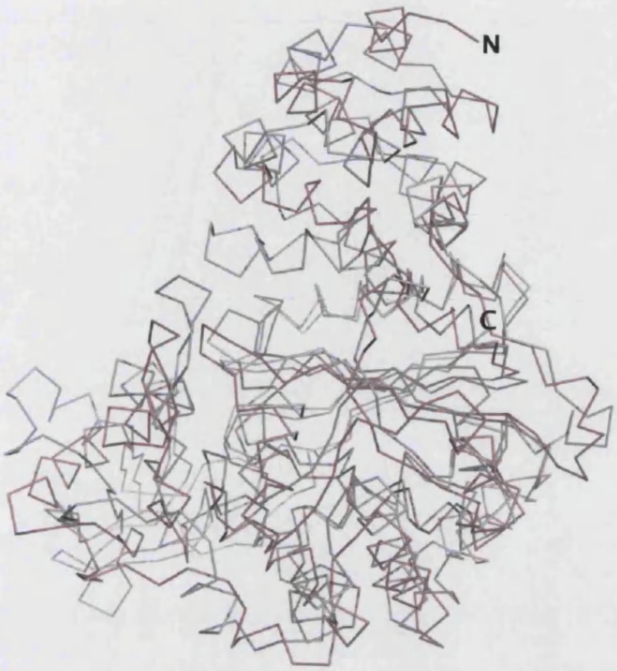
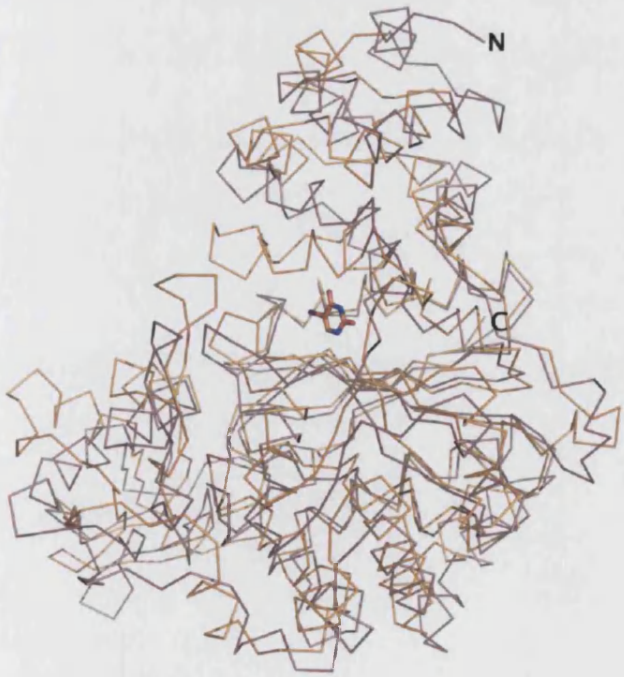


Figure 3.25: Structural alignment of native hTP (in grey) against unbound EcTP (in pink) Differences between the aligned structures are mainly observed in the α helical domain. The mixed α/β domain appears to align better and fewer differences on this domain can be identified amongst the aligned structures.



Structural alignment of hTP-5IUR complex (in orange) against unbound EcTP (in pink). Shown in the active site cleft of hTP is 5IUR. The α helical domain of EcTP appears to differ from that of hTP, as previously seen for the alignment hTP and EcTP structures.

Structural alignment of native hTP against BsPyNP (PDB ID: 1BRW) yielded an rmsd of 1.4Å, same as that obtained for the structural alignment of hTP-5IUR against BsPyNP (1.4Å; Figure 3.26). The results obtained for all structural alignments are given in Table 3.4.

Table 3.4: Structural alignment results for the three PyNPs.

Structures Aligned	**Seq. Identity (%)** (Full length seq. alignment)	**RMSD (Å)**
hTP-5IUR vs EcTP-thymine	38	1.6
Native hTP vs EcTP-thymine		1.5
hTP-5IUR vs EcTP		1.7
Native hTP vs EcTP		1.7
hTP-5IUR vs BsPyNP-uracil	40	1.4
Native hTP vs BsPyNP-uracil		1.4

Structural alignment of hTP, EcTP and BsPyNP were calculated for the backbone C^{α} atoms only. The program used for structural alignment and rmsd calculation was SPDBV (Guex and Peitsch, 1997).

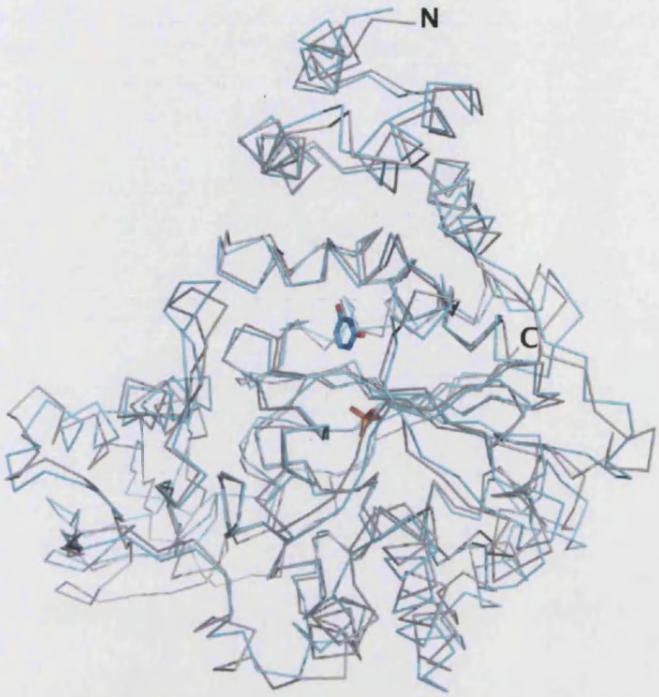


Figure 3.26: Structural alignment of native hTP (in grey) against BsPyNP (in cyan). The two structures appear to align well, with few only differences. Both α helical domain and the mixed α/β domain superimpose well indicating that the two structures have a few only differences. Shown in cyan in the active site cleft is uracil, present in the structure of BsPyNP. A phosphate ion, bound in the phosphate binding site can also be observed for the BsPyNP structure.

Structural alignment of hTP-5IUR complex (in orange) against BsPyNP (in cyan). Shown in the active site cleft is 5IUR of the hTP-5IUR complex, aligned with uracil, present in the structure of BsPyNP. Bound in the phosphate binding site is a phosphate ion, present in the structure of BsPyNP. The two molecules bound in the active site of hTP and BsPyNP appear to align well in contrast to the alignment of hTP-5IUR against EcTP-thymine complexes.

3.7.6. Conclusion

Structural analysis of the two forms of hTP, native and in complex with 5IUR have provided some new insights for the enzyme. The lack of detectable domain movement might signify that it does not actually occur in the case of hTP. It is possible that the 'closed' state of the enzyme is favoured by crystal packing, and binding of phosphate or substrates is not required for the enzyme to acquire this particular conformation. Based on the findings from the two crystal structures of hTP, it is not possible to determine whether any intermediate stages exist during catalysis and what structural differences those might have with the bound and unbound states of hTP.

Comparison of bound and unbound hTP with previously determined structures of EcTP and BsPyNP revealed that although most of the α helical domain is conserved amongst the PyNPs in sequence level, structurally the mixed α/β domain aligns better. This could either be due to sequence differences (bacterial PyNPs are smaller to hTP) or due to the organisation of the domains in the environment of the crystal (crystal packing). It is however likely that the differences observed for bound and unbound EcTP are indicative of domain movement, which has not been identified yet in its human homologue.

It is likely that intermediate stages, for which structures have not been determined yet, will provide more information not only for the catalytic activity but for the conformational changes that might be essential for the reaction to occur. Identification of intermediate stages, both functional and structural, could be used for the design and development of new drugs: if intermediate stages are required for the completion of the hTP-catalysed reaction, drugs that target such intermediate stages could be designed. If however no domain movement is essential and the enzyme acquires the 'closed' conformation even when inactive, it is of key importance to design specific drugs that target the overexpression of hTP.

B. Functional Analysis

3.8 Materials and Methods

Cloning and initial expression and purification work on native hTP was carried out by Dr. Stephen H. Prior.

3.8.1 Construction of recombinant native human thymidine phosphorylase

The original cDNA for hTP was obtained from Mammalian Gene Collection (MGC) and was in pMCV-SPORT6 vector. Analysis of its sequence with SignalP revealed that hTP does not have a signal peptide. The coding DNA sequence, (residues 38 to 482) was amplified in the presence of 0.5M betaine (primers used are given in Table 3.5). The fragment was then digested by T4 polymerase in the presence of dTTP before being annealed into a predigested pYSBLIC[1] vector (Bosnor *et al.*, 2006) with BseRI (NEB) and then transformed into NovaBlue (Novagen) cells. The DNA was then extracted using the SV Wizard (Promega) and transformed into BL21(DE3)codon$^+$ expression cells (Novagen), used to overcome codon bias, since hTP contains several rare codons as determined by Graphical Codon Usage Analyser (GCUA; Fuhrmann *et al.*, 2004). Vector integrity was then analysed by sequencing (MWG Biotech).

3.8.2 Construction of recombinant human thymidine phosphorylase mutants

3.8.2.1 Rationale for mutational work

Structural analyses on members of the PyNP superfamily lead to the identification of residues that are conserved among species. Conserved residues of particular interest are those involved in the binding and stabilisation of the substrate/inhibitor in the active site cleft. Using the structural information gained from the hTP-5IUR complex, point

[1] pYSBLIC is a modified pET-28a vector where the thrombin cleavage site has been removed to allow for ligation independent cloning (LIC). A non-cleavable His-tag is annealed to the N-terminus of the expressed protein.

mutations were designed for some of the active site residues; Lys115, His116, Tyr199, Arg202, Ile214 and Ser217 (Figure 3.17, page75; Figure 3.18, page 79; Figure 3.27).

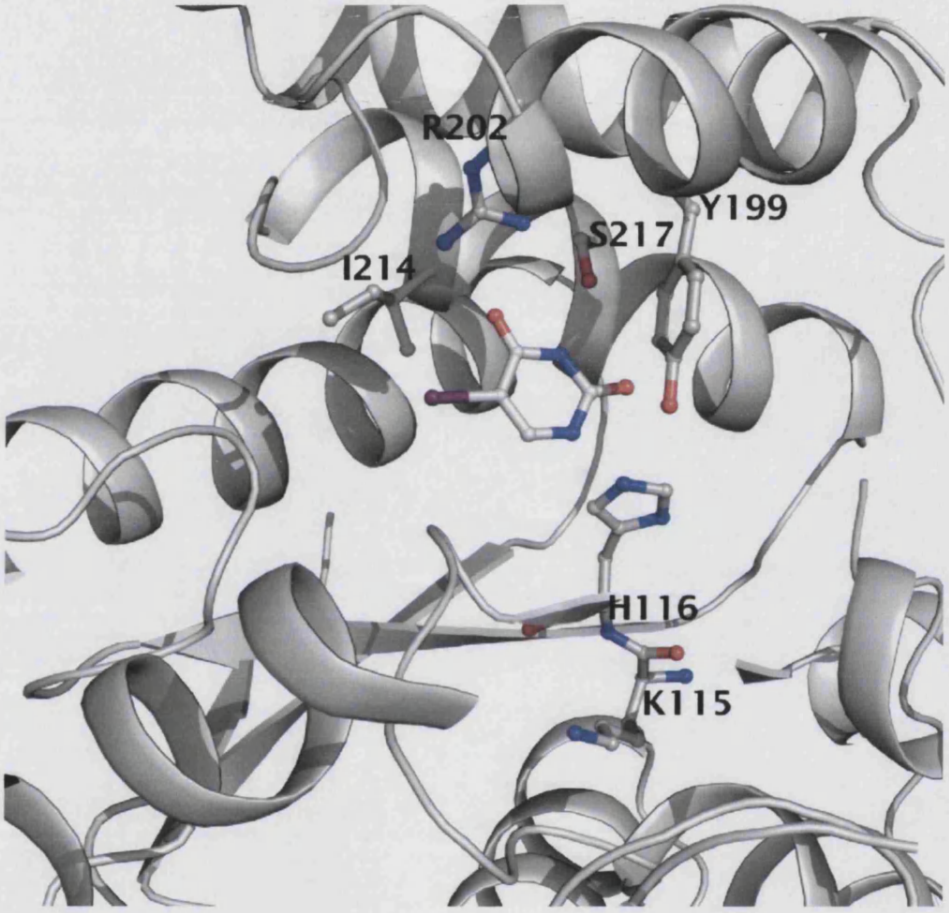


Figure 3.27: Active site of hTP where the residues under study (Lys115, His116, Tyr199, Arg202, Ile214 and Ser217) are illustrated as ball-and-stick. Also shown in ball-and-stick is the inhibitor 5IUR, present in the hTP-5IUR complex.

Although Lys115 does not form any hydrogen bonds with 5IUR in the hTP-5IUR complex structure, it appears to be conserved in all PyNPs (Figure 3.18). Previous studies (Edwards 2006) have proposed a role for Lys115 in interacting with phosphate, and along with three other residues (Ser126, Ser144 and Ser117) stabilising it in the phosphate binding site. Site-directed mutagenesis on Lys115, where it was mutated to glutamic acid,

(Miyadera *et al.*, 1995) resulted in abolishing the enzymatic and angiogenic activity of hTP. Hence, Lys115 appears to be essential for maintaining the enzymatic activity of hTP and further mutational work on this residue would 'explain' its significance in the phosphorolytic activity of hTP. Introducing thus a negative charge (mutant K115E) would indicate whether the positive charge is absolutely necessary at that particular position, while introducing a hydrophobic residue (mutation K115A) would indicate whether charge and size of Lys115 side chain are both essential for the enzymatic activity.

His116 is a residue of key importance in PyNPs: previous studies on the equivalent residue of EcTP (His85) by means of targeted molecular dynamics, suggested that this highly conserved residue is important for the catalytic mechanism of TP (Mendieta *et al.*, 2004). The role attributed to His116 (His85 in EcTP) has also been discussed and confirmed by Rick and co-workers, based on findings from experiments of quantum mechanical calculations (Rick *et al.*, 1999). More recently, Norman and co-workers (Norman *et al.*, 2004) proposed the participation of His116 in a proton shuttling mechanism in hTP, where Asp114, Glu225 and Lys222 form a triad delivering a proton to His116, which then interacts with TPI, as seen in the hTP-TPI complex structure (Figure 3.17; Norman *et al.*, 2004). In order to understand the significance of His116 towards the enzymatic activity of hTP, mutations to phenylalanine and lysine were introduced. H116F mutant retains the ring side chain of histidine but loses the charged groups, while H116K retains the positive charge but loses the ring structure.

In contrast to other PyNPs, hTP possesses only one tyrosine residue in the whole amino acid sequence of the enzyme. The tyrosine residue in hTP is located at the active site cleft and is involved in electrostatic interactions with 5IUR (Figure 3.18; *3.7.3*, Table 3.3). The role of this residue is yet unclear, although it has been suggested that modification of this residue might interfere with substrate binding or catalysis (Finnis *et al.*, 1993). Walter and co-workers (Walter *et al.*, 1990) have shown that a tyrosine residue (Tyr168 in EcTP) is implicated in hydrogen bonding interactions with thymidine, as seen from a Connolly surface diagram. Hence, point mutations were also carried out on Tyr199 (hTP)

in order to substantiate its role in the catalytic mechanism of hTP. Mutation to phenylalanine retains the aromatic character of tyrosine, losing only the hydroxyl group. Mutation to leucine introduces a bulky, hydrophobic side chain, while mutation to alanine removes both charge and size of the side chain, completely changing the environment previously present around that particular amino acid.

Involved in the network of interactions that stabilise 5IUR in the active site cleft are residues Arg202 and Ser217 (Figure 3.17). The two residues have been previously described to interact with molecules bound in the active site of hTP, such as TPI and thymine (Norman *et al.*, 2004; El Omari *et al.*, 2006). The significance of Arg202 for the enzymatic activity of hTP has been studied by mutational work where Arg202 was mutated to glutamate. Results indicated abolishing of both angiogenic and enzymatic activity of hTP (Miyadera *et al.*, 1995). It is not however clear if mutation to glutamate (and introduction of a negative charge at this position) causes this effect or if Arg202 is absolutely essential for enzymatic activity. For this purpose R202E, which introduces a negative charge and R202S, which removes the lengthy side chain and the basic character of arginine at that position, were examined. On the other hand, nothing has been reported to date for Ser217, and its significance for the enzymatic activity of hTP. Hence, mutation S217G would indicate whether the hydroxyl group of serine at that particular position is essential. These mutations would provide a better understanding of the significance and role of Arg202 and Ser217 in the catalytic activity of hTP.

Apart from the residues that interact *via* hydrogen bonds with 5IUR, a hydrophobic pocket appears to 'hold' the iodine atom (Figure 3.18). This has been previously described for hTP-TPI complex (Norman *et al.*, 2004), where the inhibitor bound in the active site cleft had a chlorine atom, stabilised in the same pocket. Mutational work on members of the hydrophobic pocket has only been reported for L148R, where biological work indicated complete loss of the enzymatic and angiogenic activity of hTP (Miyadera *et al.*, 1995). Mutational work on another residue of the hydrophobic pocket, Ile214, would thus indicate whether the hydrophobic pocket is a significant component of the active site cleft, essential for stabilising bound molecules so as for hTP to be

enzymatically active. Mutant I214A completely removes the bulky side chain of leucine and would be indicative of whether the size of the leucine side chain is essential in that particular position, for retaining van den Waals interactions, as part of the hydrophobic pocket.

3.8.2.2 Materials and methods

Primers for the desired mutations (K115E, K115A, H116F, H116K, H116A, Y199A, Y199L, Y199F, Y199K, R202E, R202S, S217G and I214A) were designed with Primer X (Lapid, 2003) and were ordered from Invitrogen (Table 3.5). The coding sequence of native hTP (already present in pYSBLIC[2]; Bosnor *et al.,* 2006) was used as a template for PCR amplification in the presence of the primers for the desired mutation (forward and reverse; Table 3.5) and 0.5M betaine. The polymerase chain reaction (PCR) was performed using KOD polymerase (Novagen) and the Boehringer Expand^{TM} PCR system. The protocol used for all PCR reactions is the following: 94°C for 2min, 18 cycles of 94°C for 30sec (melting) followed by 55°C for 30sec (annealing) and 68°C for 8min (extension), 68°C for 7min and storage at 4°C. The PCR product was visualised on a 0.1% agarose gel and was then digested with 1µl of DpnI (Promega) at 37°C for 1hr to remove any non-mutated plasmids. The resulting DNA was transformed in BL21(DE3)codon^{+} expression cells by heat shock and was plated out onto LB agar plates supplemented with 50µg/ml kanamycin and 50µg/ml chloramphenicol. Transformants were incubated overnight at 37°C. Single colonies were then selected from the transformation plates and used to inoculate 10ml of LB supplemented with 50µg/ml kanamycin and 50µg/ml chloramphenicol, allowed to incubate overnight at 37°C. Plasmid DNA was extracted using the mini-prep wizard (SV Wizard, Promega) and successful insertion of mutations was confirmed with T7 forward and reverse sequencing (MWG Biotech).

---

[2] pYSBLIC is a modified pET-28a vector where the thrombin cleavage site has been removed to allow for ligation independent cloning (LIC). A non-cleavable His-tag is annealed to the N-terminus of the expressed protein.

Table 3.5 : Primers for recombinant native hTP and hTP mutants

(forward, reverse)

Native hTP	5'-CACCACCACCACATGTCGCCAGAGCCCAAGCAGCTCC-3' 5'-GAGGAGAAGGAGCGTTATTGCTGCGGCGGCAGAACGAGCTCTGCG-3'
Y199F	5'- GAC GGA ATC CTA TTT GCA GCC AGA -3' 5'- ATC TCT GGC TGC AAA TAG GAT TCC GTC -3'
Y199A	5'- GAC GGA ATC CTA GCG GCA GCC AGA -3' 5'- ATC TCT GGC TGC CGC TAG GAT TCC GTC -3'
Y199L	5'- GAC GGA ATC CTA CTG GCA GCC AGA -3' 5'- ATC TCT GGC TGC CAG TAG GAT TCC GTC -3'
H116F	5'- AG CTT GTC AC AAG TTT TCC ACA GGG GGT G -3' 5'- ACC CCC TGT GGA AAA CTT GTC CAC AAG CTG -3'
H116K	5'- CAG CTT GTC AC AAG AAA TCC ACA GGG GGT G -3' 5'- ACC CCC TGT GGA TTT CTT GTC CAC AAG CTG -3'
K115E	5'- CAG CAG CTT GTG GAC GAA CAT TCC ACA GGG GG -3' 5'- CCC TGT GGA ATG TTC CTC CAC AAG CTG CTG -3'
K115A	5'- CAG CAG CTT GTG GAC GCG CAT TCC ACA GGG GG -3' 5'- CCC TGT GGA ATG CGC CTC CAC AAG CTG CTG -3'
R202E	5'- ATC CTA TAT GCA GCC GAA GAT GTG ACA GCC ACC -3' 5'- TGG CTG TCA CAT CTT CGG CTG CAT ATA GGA -3'
R202S	5'- ATC CTA TAT GCA GCC AGC GAT GTG ACA GCC ACC -3' 5'- TGG CTG TCA CAT CGC CGG CTG CAT ATA GGA -3'
I214A	5'- CTG CCA CTC GCG ACA GCC TCC ATT CTC AGT AAG -3' 5'- GAG AAT GGA GGC TGT CGC GAG TGG CAG GC -3'
S217G	5'- CCA CTC ATC ACA GCC GGC ATT CTC AGT AAG AAA -3' 5'- TTC TTA CTG AGA ATG CCG GCT GTG ATG AGT GGC -3'
H116A	5'- AG CTT GTC AC AAG GCG TCC ACA GGG GGT G -3' 5'- ACC CCC TGT GGA CGC CTT GTC CAC AAG CTG -3'
Y199K	5'- GAC GGA ATC CTA AAA GCA GCC AGA -3' 5'- ATC TCT GGC TGC TTT TAG GAT TCC GTC -3'

The altered nucleotides for the incorporation of each mutation are underlined. The two mutations shown in italics (H116A and Y199K) were unsuccessful.

3.8.3 Expression of native hTP and hTP mutants

Cells harbouring the recombinant plasmid of native hTP were used to inoculate overnight cultures of LB/TB (Sambrook *et al.*, 1989) media, supplemented with kanamycin (34μg/ml) and chloramphenicol (50μg/ml). Initial trials for the expression of native hTP were carried out in 2L culture flasks of TB media supplemented with K_2HPO_4 and KH_2PO_4 salts and chloramphenicol (50mg/ml) and kanamycin (34mg/ml). Cultures were incubated at 37°C for approximately 4hrs before the cell density at $O.D._{600}$ reached 0.6-0.8, at which point protein expression was induced by the addition of 1mM isopropyl β-D-thiogalactoside (IPTG-Bioline). Cells were incubated three hours post-induction and then harvested by centrifugation at 7,000rpm for 10min. Cell pellets were stored at -20°C until further use. Due to the low yield however, optimisation of the expression was necessary. For this purpose large-scale expression of the enzyme in a fermentor (Bioflo 3000-New Brunswick Scientific) was carried out (details for preparing and connecting the fermentor are given in Appendix I). Conditions such as temperature, pH and agitation were controlled during fermentation. Two temperatures were tested for optimal hTP expression: 37°C and 30°C. Cell density was monitored and when the $O.D._{600}$ reached 0.6-0.8, protein expression was induced by the addition of 1mM isopropyl β-D-thiogalactoside (IPTG, Bioline). To identify the optimal incubation time after induction cultures were monitored over a period of 24hrs, at which point cell death was easily identified. Cells were harvested by centrifugation at 7,000rpm for 10min.

Cells harbouring the recombinant plasmid of mutants were used to inoculate overnight cultures of LB/TB (Sambrook *et al.*, 1989) media, supplemented with kanamycin (34μg/ml) and chloramphenicol (50μg/ml). Protein expression for the mutants was carried out in 2L culture flasks. When cell density at $O.D._{600}$ reached 0.6-0.8, protein expression was induced by the addition of 1mM isopropyl β-D-thiogalactoside (IPTG-Bioline). Cells were incubated three hours post-induction and then harvested by centrifugation at 7,000rpm for 10min. Cell pellets were stored at -20°C until further use.

3.8.4 Cell lysis of native hTP and hTP mutants.

Cells were initially resuspended in lysis buffer containing 20mM Na-citrate pH5.7, 100mM NaCl and one tablet Compl@te EDTA-free protease inhibitor (one tablet per 25ml of solution, as recommended by the supplier; Roche). Lysozyme (SIGMA) was initially added to this suspension which was further incubated at 4°C for 15min. Addition of lysozyme however appeared to reduce protein viability and was removed in further cell lysis steps. Two types of cell lysis were tested: French press and sonication, using an MSE Soniprep 150 sonicator with 20sec pulses. The lysed cells were subjected to centrifugation at 24,000 rpm for 1hr at 4°C and the cell lysate was filtered using sterile 0.45μm filter units (Millipore).

3.8.5 Protein purification of native hTP and hTP mutants

The filtered cell lysate (native or mutant hTP) was loaded onto a 5ml Ni^{2+} His-select column (Sigma) previously equilibrated with buffer A (20mM Na-citrate pH5.7, 100mM NaCl, one tablet Compl@te EDTA-free protease inhibitor; one tablet per 25ml of solution, as recommended by the supplier; Roche). Sample was loaded and the column was washed with buffer A. Protein was eluted using an imidazole gradient (0-250mM Imidazole) by mixing buffers A and B (buffer B: 20mM Na-citrate pH5.7, 100mM NaCl, one tablet Compl@te EDTA-free protease inhibitor, 250mM Imidazole). Fractions were analysed on a 15% SDS-PAGE and those containing significant amounts of hTP were pooled.

Further purification was carried out by size exclusion chromatography using a Superdex 200 column (GE Healthcare) pre-equilibrated with buffer C (20mM Na-citrate pH5.7, 100mM NaCl). A protocol for multiple injections (5ml – 5runs) was set-up in order to facilitate the purification procedure. Sample loading was carried out at a flow rate of 0.5mg/ml and the column was washed with 1.1 column volumes of buffer C, for the completion of each run (each run consisted of 1ml sample injection). Eluates were analysed on a 15% SDS-PAGE to check purity of the protein.

3.8.6 Protein concentration and quantitation

Fractions containing pure protein were pooled together and concentrated. Concentration of the protein sample was performed using Amicon Ultra-15 (30KD cut-off; Millipore). Concentration was carried out in 5min cycles of centrifugation at 4,000g. The initial pool of ~10ml was concentrated down to ~1ml, at which point the concentration of flow through and supernatant (flow through – impurities, supernatant – purified protein) was assayed, using the BCA assay method (standard BCA Assay, Pierce). Absorbance was measured at 562nm. Usually no additional concentration cycles were required. Analysis of the concentrated protein samples was performed by SDS-PAGE.

An additional step, carried out only for native hTP was that of trichloroacetic acid precipitation (TCA precipitation). TCA is a potent dehydrator, and as such it removes water that is normally found in the protein. As a result the protein aggregates and subsequent analysis with SDS-PAGE allows the easy identification of proteins. This is particularly useful as it allows for even a low amount of protein (such as impurities) to be detected in a protein sample. 20% TCA was added to an equal volume of the protein sample to be examined and the sample was incubated for 30min on ice. Centrifugation for 15min at 4°C resulted in the formation of a pellet containing the precipitated proteins. The supernatant was discarded and ~300μl of cold acetone were added to the pellet. Centrifugation for 5min at 4°C and subsequent removal of the supernatant were followed by drying of the pellet for 10-20min. The pellet was resuspended in SDS-PAGE loading dye and analysis was carried out by SDS-PAGE.

3.8.7 Identifying hTP – Western blot analysis

hTP was expressed as a fusion protein with a His-tag on its N-terminal domain. This property was used for Western blot analysis. Protein and marker samples were initially analysed by SDS-PAGE. In order to make the protein sample accessible to antibody detection, it was transferred from the gel to a membrane, made of Polyvinylidene Difluoride (PVDF). Protein binding on the membrane is based on hydrophobic and charged interactions between the membrane and protein. The effectiveness of transfer of

the protein samples from the gel to the membrane was checked by staining the membrane with Ponceau S dye.

To avoid non-specific binding the membrane was washed in a dilute solution of non-fat dry milk, containing 0.1% of Tween 20 detergent (blocking). After blocking, the primary antibody, (mouse anti-his antibody) at a final concentration of 1:2,000 was incubated with the membrane overnight, under gentle agitation.

The membrane was rinsed to remove unbound primary antibody and incubated with the secondary antibody (rabbit anti-mouse anti-his antibody) at a final concentration of 1:1,000. The secondary antibody was linked to horseradish peroxidase in conjunction with a chemiluminescent agent. This allowed for the detection of the product as it produced luminescence, in proportion to the amount of protein. Incubation with secondary antibody was carried over an hour, in dark. A photographic film was then placed against the membrane, and exposure to the light from the reaction created an image of the antibodies bound to the blot.

3.8.8 Reversed phase chromatography and Mass Spectrometry

Purification steps for hTP did not normally include reversed phase chromatography. This was a step necessary for sample preparation for Mass spectrometry. Purified protein was diluted to 3mg/ml (total sample volume of 1ml) prior to being loaded on the reversed phase column. A gradient was set up from 30 to 80% buffer B (acetonitrile) and eluted fractions were collected over a total period of 1hr. Fractions containing the protein were pooled together and freeze-dried prior to lyophilisation. A pool sample was also run on SDS-PAGE to check for any protein degradation.

Mass spectrometry was carried out on hTP in order to confirm the protein and whether any impurities or protein degradation was present in solution. Initial experiments were carried out using the method of ElectroSpray Ionisation – Time of Flight Mass spectroscopy (ESI-TOF MS), using a microTOF electrospray (Bruker Daltonik GmbH, Bremen, Germany). This was coupled to a syringe pump for the introduction of sample at 4µl/min. The protein sample was in buffer consisting of 20mM Na-citrate and 100mM

NaCl and the protein concentration used was 1mg/ml. Data was acquired over 5min and averaged to obtain the spectrum recorded. The nebulising gas used was nitrogen, applied at a pressure of 0.4 bar. The drying gas was also nitrogen, supplied at a flow rate of 4 l/min and a temperature of 200°C. The ion mode that was used for this experiment was positive ion mode. Data acquired was processed *via* Compass OpenAccess 1.2 Software (Bruker Daltonik GmbH, Bremen, Germany). Further experiments using ESI-TOF were set up for protein sample in HPLC grade water, with and without formic acid. Spectra obtained from this method indicated that the use of a different mass spectrometry method might prove to be more suitable for this protein sample. Hence Matrix Assisted Laser Disorption/ Ionisation – Time of Flight (MALDI-TOF) was then carried out.

MALDI analyses were performed on the Voyager (Applied Biosystems) MALDI-TOF mass spectrometer with linear and reflectron analysers. Solutions of protein sample (protein concentration of 1mg/ml in HPLC grade water) and matrix were mixed and 0.5-1 µl was pipetted onto a sample plate. The sample spot was dried, allowing the crystallisation of the mixture and then irradiated with a pulsed nitrogen laser (wavelength of 337nm at 20hZ). The matrix was ablated from the plate while the sample was simultaneously desorbed and ionised. Following ionisation, the sample was accelerated into a flight tube (typically 20kV).

3.8.9 hTP affinity for heparin

A 1ml heparin column (SIGMA) was used to test the affinity of hTP for heparin. The column was initially equilibrated with buffer A (20mM Na-citrate and 100mM NaCl at pH 7). A total of 11ml of 1mg/ml protein sample was loaded onto the pre-equilibrated heparin affinity column, at a flow rate of 0.5ml/min. The same buffer was used for washing the column prior to elution. Elution was carried out using buffer B (2M NaCl). Eluted fractions were collected and analysed on a 15% SDS-PAGE.

3.8.10 Crystallisation trials for native hTP and hTP mutants

For crystallisation of native hTP and hTP mutants the Hampton structure screens I and II (Hampton Research) were performed, using the hanging drop vapour diffusion method.

The protein concentration used initially was 7mg/ml and incubation was carried out at 16°C. Initial hits obtained from the structure screens were further optimised. Crystallisation under oil as well as incubation with additives was carried out as part of the optimisation trials. Techniques such as micro-seeding and streak seeding were employed to improve the quality of the crystals obtained, as well as cross-linking, using glutaraldehyde. A 0.1% of glutaraldehyde was added to the mother liquor and drops were monitored over a period of three hours. Dehydration experiments where the drop was exposed to air for 5 to 20sec were also carried out.

3.8.11 Enzymatic activity assays

The enzymatic activity of native hTP and hTP mutants was assayed by the spectophotometric method, a variation of the method firstly descried by Friedkin and Roberts in 1954 (Friedkin and Roberts, 1954; Yoshimura *et al.*, 1990), which takes advantage of the absorbance difference between thymidine and thymine at 300nm. The reaction solution for the enzymatic activity assays, required several steps of optimisation, as hTP was quite unstable in several of the previously used conditions. Optimised reaction conditions contained 0.1M MES pH 5.5 as the buffer of the reaction, 0.1M KH_2PO_4 pH 5.5 as a source of phosphate, essential for the phosphorolytic activity of hTP and thymidine, the substrate of the reaction in a total reaction volume of 100µl. The pH of MES and KH_2PO_4 was adjusted to 5.5, based on previous experiments which examined the pH profile of recombinant hTP in phosphorolysis reactions and indicated that hTP phosphorolytic activity is most optimum at pH 5.0-6.0 (Finnis *et al.*, 1993). The substrate, thymidine, was ordered from Sigma and was diluted with HPLC grade water to prepare the stock solution, from which varying substrate concentrations (100µM to 10mM) were used to identify the optimum range for the reactions. Purified native hTP of various concentrations (2ng/ml to 1µg/ml) was used in assays in order to identify the lowest enzyme concentration that could be used without jeopardising the validity of the results. The most optimum enzyme concentration was found to be 2ng/µl, which is equivalent to 40nM, as the molecular weight of hTP is ~50kDa. All experiments, for native hTP and hTP mutants were therefore carried out using the same enzyme concentration for consistency purposes. All reactions were prepared in triplicates in order

to ensure the validity of the results and all experiments were carried out in duplicates in order to ensure the accuracy of the results obtained. Reactions were incubated at 37°C and stopped by the addition of 900µl of 0.2M NaOH. Blanks were also prepared and treated in the same way as the reaction samples; all blanks included the reaction mixture, lacking only the enzyme, in order to ensure that the only readings would correspond to thymine produced by the phosphorolytic reaction of hTP. The absorbance was measured at 300nm, where the difference in the extinction coefficient between thymine and thymidine is $3400M^{-1}cm^{-1}$ (Spinazzola *et al.*, 2002). The extinction coefficient is required for the conversion of the absorbance readings to concentration (Beer's Law: $A=\varepsilon*c*l$). These values were then used for the calculation of the initial velocity (V_0) taking into consideration the time points at which each sample was collected and the reaction was stopped. The initial velocities were then used as the input for further analysis using the GraFit Erithacus software (Erithacus software; Leatherbarrow, 2001), where addition of the Michaelis-Menten equation and of the protein concentration allows the calculation of K_m, V_{max} and standard error values. V_{max} values obtained by this analysis were then used to calculate k_{cat} using the formula $k_{cat}=V_{max}/E$, where E is the molarity of the enzyme. (Enzyme kinetics theory is provided in Appendix II).

The enzymatic activity assay was also used to measure the inhibition potency of 5iodouracil. 5IUR (see *3.7*) was crystallised with hTP as a mimic of 5FU, a currently used chemotherapeutic agent, with the added advantage that it is devoid of the fluorine atom. As discussed above, 5FU can cause a cumulative effect of toxicity, which with 5IUR would be significantly reduced, given that 5IUR would also be an effective inhibitor of hTP. 5IUR was purchased from Sigma and was dissolved in NaOH. A stock concentration of 10mM was prepared, of which varying concentrations (0.005, 0.01, 0.05, 0.1, 0.15 and 0.2mM) were tested for its efficacy in inhibiting hTP. The reaction assays were prepared as those previously described for native hTP and hTP mutants, with the addition of inhibitor. Inhibitor was also added in all the blanks, which were only devoid of enzyme, in order to blank for any of the substances that could give background absorbance. As previously described, reactions were incubated at 37°C and stopped by the addition of 0.2M NaOH. Absorbance readings were converted to concentration, using

the extinction coefficient of thymine. Once the initial velocity was calculated for the reactions that contained inhibitor, results were used for generating Dixon plots. Dixon plots (Dixon, 1953) are an efficient way of determining the type of inhibition, as well as the K_i for the inhibitor.

3.8.12 Modeling of mutants on the hTP-5IUR crystal structure

In order to identify the effect that the mutations might have on the active site environment of hTP all point mutations were modeled in the hTP-5IUR structure. Mutations were introduced using Coot (Emsley *et al.*, 2004) and models were locally refined to ensure that the mutated structure was devoid of steric hindrances. To have a directly comparable result, the structure of hTP-5IUR was also taken through the same treatment prior to analysis. Analysis was performed by calculation of the number of van der Waals contacts using the program CONTACT by CCP4 (CCP4, 1994) for each mutated residue. The effect of the mutations on the hydrogen bond network was estimated by the calculation of hydrogen bonds for each of the mutated active site residues, using the program HBPLUS (McDonald *et al.*, 1994).

3.9 Results and discussion

3.9.1 Construction of hTP mutants

Successful incorporation of the mutations was confirmed by DNA sequencing for each of the mutants. Although the same protocol was used for all mutations, H116A and Y199K were not successfully amplified. Analysis of the PCR product revealed no DNA for these two samples, in contrast to the results obtained for all other mutations. Optimisation of the conditions used for PCR was carried out without any success for H116A and Y199K (Table 3.5).

3.9.2 Expression and purification of native hTP and hTP mutants

Although an initial protocol for expression and purification of native hTP was available, the yield obtained was low and optimisation of the initial conditions was required. Auto-induction media (expression is induced automatically in late log-phase growth due to the depletion of carbon sources) were tested for hTP expression, producing however an even lower yield of protein to that obtained in TB media. Large scale production of native hTP was finally achieved using the BioFlo fermentor vessel (see Appendix I).

As no protocol was available for this type of expression, expression trials were set-up where conditions such as temperature, pH, dissolved oxygen (supplied to the vessel during bacterial cultures), as well as post-induction incubation time were tested. Addition of antifoam to bacterial cultures was tested as it reduces the formation of foam produced during fermentation. This however can prove to be inhibitory for cell growth when added in excess, so optimisation of this step was also necessary.

Optimum conditions for native hTP expression were temperature of 30°C and pH of 6.5. The agitation of the cell culture was optimum when set in-loop with the dissolved oxygen (DO). More specifically, when DO value dropped below a user-specified value (35.0) the agitation increased (minimum of 50rpm, maximum of 300rpm), to allow for more oxygen to be supplied to the vessel. Using a loop mode for DO and agitation allowed the 'automated' control of the oxygen supplied and no manual input was required for this parameter (see Appendix I). The bacterial cell growth reached a 'plateau' three hours

after induction, while further incubation appeared to result to cell death. The use of antifoam was limited to μl amounts of 10% antifoam solution (SIGMA). In order to reduce the formation of foam, media were prepared in such amounts as for the impellers to be submerged in liquid (approximately 3.5L of media). Large-scale expression using the optimum conditions for hTP yielded approximately 700μl of pure protein (concentration of ~9mg/ml), as opposed to 200μl of protein (concentration 6-7mg/ml) typically obtained by 2L flask cultures.

Expression of hTP mutants was less problematic. The use of the initial expression protocol indicated that only a few modifications were required before the yield was satisfactory. Expression for all mutants was carried out in TB media, supplied with glycerol, KH_2PO_4 and K_2HPO_4 salts and chloramphenicol (50mg/ml) and kanamycin (34mg/ml). Bacterial cultures were incubated at 37°C and monitored for cell growth in order to induce expression by the addition of 1mM IPTG. Although all mutations were only point mutations, they appeared to express better than native hTP (this might be due to the loss and/or reduction of the enzymatic activity of hTP mutants when compared to that of the native enzyme; see *3.9.8*). For this reason the use of a 'more controlled' environment, using a fermentor vessel, was not required.

Purification was carried out using the same protocol for native hTP and all hTP mutants. The cell pellet obtained was lysed using sonication, although cell lysis using French press, was also tested. Sonication proved to be more efficient as French press yielded very dense cell lysate which was not easily handled in the following steps of purification. The first step of purification, using an affinity chromatography column, allowed the separation of hTP from several other contaminants which were present in the cell lysate. Once most of the impurities were removed, the eluate of the affinity chromatography step was centrifuged, in order to remove any dust or particles, and was then loaded onto a size exclusion chromatography column. Separation according to size was favourable for hTP as most of the contaminants that remained in the sample were of different molecular weight. The two steps used in the purification process of native hTP and hTP mutants were enough to remove any impurities present in the cell lysate. SDS-PAGE analysis for

each step of the purification procedure confirmed the presence of protein and the removal of impurities present in the initial sample (Figure 3.28).

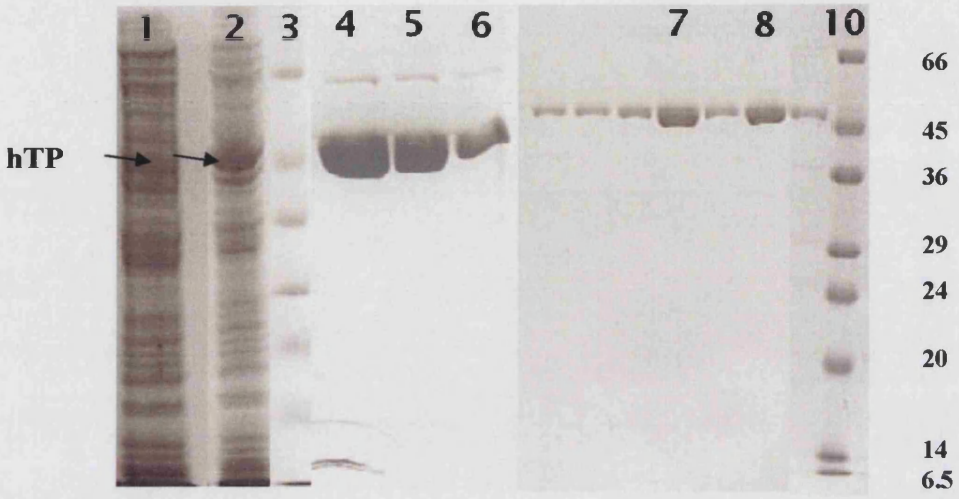
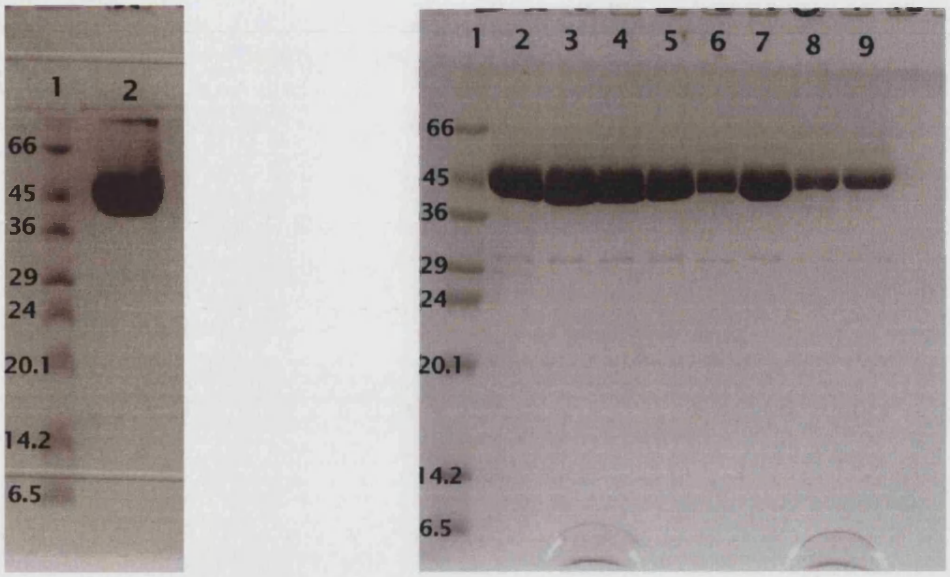


Figure 3.28: **A**. Expression and purification of native hTP. Lanes 1 and 2 contain expression samples, pre and post induction respectively. Lanes 4 to 6 contain fractions after the first purification step (affinity column). Lanes 7 and 8 contain samples after the second purification step, where the protein is pure and all impurities have been removed. Lanes 3 and 10 contain protein markers (low molecular weight marker, SIGMA; in kDa range).

3.9.3 Protein concentration and quantitation

Protein concentration presents a risk as it is possible for the protein sample to bind the membrane used. In order to limit this risk the membranes used for concentration were washed with buffer, same as that which contained the protein sample (20mM Na-citrate pH5.7 and 100mM NaCl). The use of 30kDa molecular weight cut-off filters allowed further removal of contaminants (if present) of smaller molecular weight from the eluate of the size exclusion chromatography step (Figure 3.29).



A B

Figure 3.29: Figure **A** illustrates the SDS-PAGE analysis of native hTP, purified and concentrated. Lane 1 contains the marker (low molecular weight marker, SIGMA; in kDa range) and lane 2 contains the protein sample. Figure **B** illustrates the SDS-PAGE analysis of purified and concentrated mutants of hTP. Lane 1 – marker (low molecular weight marker, SIGMA; in kDa range), lane 2 – mutant K115E, lane 3 – mutant K115A, lane 4 – mutant K115A (different purification sample), lane 5 – mutant H116F, lane 6 – mutant H116K, lane 7 – mutant Y199A, lane 8 – mutant Y199L and lane 9 – mutant Y199F.

In order to ensure that no impurities were present in the protein sample an additional step was carried out after the completion of protein concentration. This involved the precipitation of ~20µl of the concentrated sample with TCA. This was only carried out for native hTP, in order to get an understanding of how accurate the results obtained from SDS-PAGE were. A standard TCA precipitation protocol was carried out and the protein samples were analysed by SDS-PAGE. If any remaining impurities were present in the protein sample, they would show up clearly after TCA precipitation. The results obtained

from this step indicated that a significant amount of hTP was present in the sample tested while no significant presence of impurities was detected (Figure 3.30).

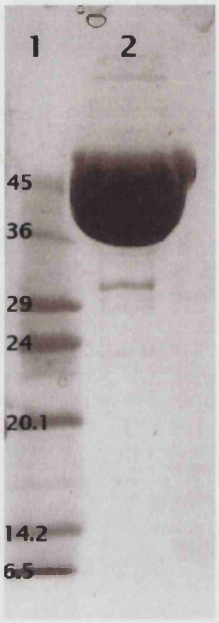


Figure 3.30: SDS-PAGE analysis of native hTP after TCA precipitation. Lane 1 contains the marker (low molecular weight marker, SIGMA; in kDa range) and lane 2 contains the protein sample after TCA precipitation. It is evident that only hTP is present in the sample. No impurities or degradation products are detected.

The protein concentration was calculated after the concentration steps, using the BCA method. Initial methods for the calculation of the concentration of native hTP included the BSA method and the UV reading of the protein sample at a wavelength of 280nm. The results however obtained from both these methods did not agree with the observation made based on the concentrated samples which were analysed with SDS-PAGE: very low concentrations were calculated (using both BSA and UV methods) for the protein sample which was in contrast to the bright, single bands observed on the SDS-PAGE for hTP. The use of BCA method was chosen based on its agreement with the results observed on the SDS-PAGE for the concentrated samples. The concentration of native hTP was calculated to be ~8.8mg/ml and that of mutants ranged from 7.5mg/ml to ~20mg/ml.

3.9.4 Identifying hTP – Western blot analysis

Western blot analysis was performed to confirm expression and purification of hTP since the only alternative identification tool was SDS-PAGE analysis. The presence of a histidine tail on the N-terminal domain of the protein made the western blot analysis easier. A sample of pure, native hTP, as assessed by SDS-PAGE analysis, was used for western blot. Development of the photographic film allowed the detection of a single band in the correct molecular weight for hTP. This confirmed the results obtained by SDS-PAGE analysis (Figure 3.31).

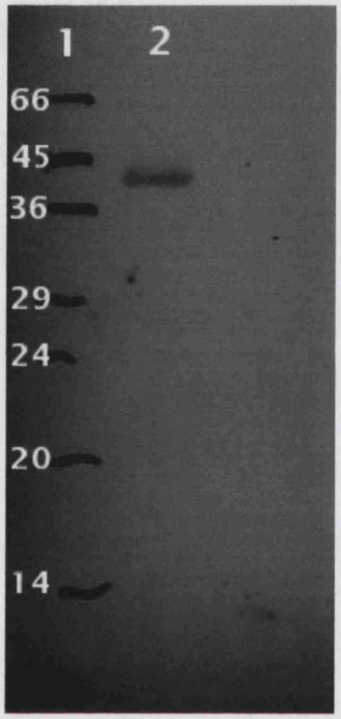


Figure 3.31: Western blot analysis for hTP. Lane 1 contains the marker (low molecular weight marker, SIGMA; in kDa range) and lane 2 contains the protein sample. It is evident from the Western blot analysis that hTP is present in the sample. (The bands marked for the marker on lane 1 are an approximation of the actual marker bands as seen on the SDS-PAGE).

3.9.5 Mass Spectrometry and Reversed phase chromatography

ESI-TOF experiments were carried out for native hTP in 20mM Na-citrate pH 5.7 and 100mM NaCl. The spectrum obtained from this experiment did not resemble that expected typically for a sample of pure protein. Instead of an 'envelope' spectrum, with peak at the correct molecular weight, the spectrum obtained indicated either protein degradation or the presence of several proteins (impurities) in the sample tested, since several peaks of different molecular weights were detected. In order to identify why this is happening and whether the buffer in which the protein sample is affects the spectrum obtained, a series of ESI-TOF experiments were carried out.

Initially a sample of BSA was used as standard and was tested in HPLC grade water, where the spectrum obtained was that of pure protein at the correct molecular weight. Introduction of formic acid in the solution prior to ESI-TOF analysis provided extra charge and improved the signal. BSA was then resuspended in 20mM Na-citrate pH 5.7 and 100mM NaCl, the buffer in which hTP was first examined. The results obtained were dramatically different. More specifically, the spectrum obtained clearly indicated that the buffer 'masks' the readings as the 'envelope' previously observed was now absent. Addition of formic acid to the solution did not improve the spectrum obtained.

This indicated that the buffer in which hTP was stable was probably not suitable for this type of analysis. Buffer exchange would probably assist in exchanging the buffer with HPLC grade water but due to the sensitivity of the protein, a different method was chosen. Reversed phase chromatography, which takes advantage of the hydrophobic interactions made between the protein sample and the gel bed of the column, was not normally used for hTP purification. It was however an alternative way to remove the buffer from the protein sample and exchange it with HPLC grade water. Hence, the eluate collected from reversed phase chromatography was dissolved in HPLC grade water and the sample was analysed by SDS-PAGE, which indicated that the protein sample was pure although a small amount of degradation had occurred during this step. The concentration of the protein sample in water was calculated to be 13.5mg/ml (Figure 3.32).

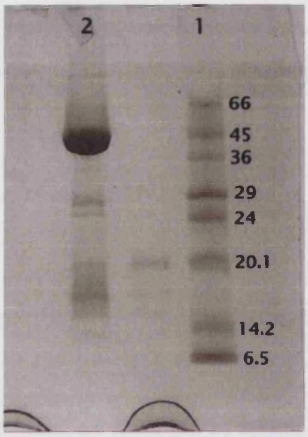


Figure 3.32: SDS-PAGE analysis of hTP, concentrated after elution from reversed phase chromatography. Lane 1 contains the marker (Low molecular weight marker, SIGMA; in kDa range) with the molecular weights corresponding to each band. Lane 2 contains the sample, where it is clear that hTP is present and concentrated. A few background bands also observed on lane 2 are probably degradation products of hTP.

Once the buffer was removed, hTP samples in HPLC grade water were again used for ESI-TOF experiments. The spectrum obtained was that of an 'envelope' although the molecular weight which was calculated from this spectrum did not correspond to hTP. More than one peaks were identified and several were of lower molecular weights. As the SDS-PAGE analysis did not show any impurities in the protein sample, this was indicative of protein degradation.

It is possible that the instability of the sample was not allowing for a good spectrum to be obtained with the method of ESI-TOF and degradation was occurring. For this purpose a second type of mass-spectrometry was tested for hTP: MALDI-TOF. This is considered to be the method of choice for protein samples as it is a soft ionisation technique, used for samples which tend to be fragile and fragment when ionised by more conventional ionisation methods. It is similar to electrospray ionisation although it causes much fewer multiply charged ions. In addition, MALDI-TOF uses a matrix to protect the protein samples from being destroyed by the direct laser beam (used for ionisation) and to facilitate vaporisation and ionisation.

The protein sample for MALDI-TOF experiments was prepared only in an aqueous solution as the buffer previously proved to 'mask' the signal. The spectrum obtained from this method resembled that of ESI-TOF: more than one peaks were detected and the molecular weight corresponding to those peaks was probably that of hTP degradation products. It was thus evident that the protein sample was not stable enough for use with either of the two mass spectrometry methods. It is not however clear what was causing the degradation of hTP and what could be done to make the protein sample stable enough for use in such experiments. Purification trials of hTP which included a reversed phase chromatography step indicated that the protein does not recover after lyophilisation. It is thus possible that a part of the protein is still in its intact form even in aqueous solution while most of it degrades to smaller fractions, as those observed in the spectra from both ESI-TOF and MALDI-TOF.

3.9.6 hTP affinity for heparin

Many angiogenic molecules have been reported to interact with heparin. The binding of heparin to many of the angiogenic growth factors, such as the vascular endothelial growth factor (VEGF; Ishihara *et al.*, 2003) and the fibroblast growth factor 2 (FGF-2; Ferrara *et al.*, 2003) has been reported to occur through their heparin binding domains. Angiogenin also binds heparin *via* its heparin binding site, of which key component is a cluster of three arginine residues (Soncin *et al.*, 1997). A similar cluster of arginine residues can be observed in the mixed α/β domain of hTP. Residues 279, 303, 369, 430, 433 and 464 appear to form a 'patch' of basic charge which would be likely to compose a heparin binding site. In order to identify whether hTP has affinity for heparin the experiment described in *3.8.9* was carried out. The protein sample loaded on the heparin affinity column would bind to heparin on the gel bed and only be removed upon elution with NaCl. This however was not the case as detected by SDS-PAGE analysis of samples collected from all steps of the process (Figure 3.33). The protein sample loaded to the column, the flow through (washing of the column) and eluted samples were analysed by SDS-PAGE analysis.

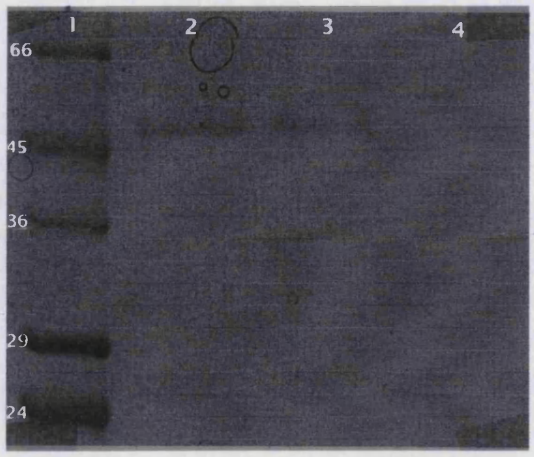


Figure 3.33: SDS-PAGE analysis of the samples loaded and eluted of the heparin affinity column. Lane 1 contains the marker (Low range molecular weight marker, SIGMA; in kDa range). Lane 2 contains the sample loaded on the heparin column, lane 3 contains the sample which was collected while washing the column and finally lane 4 contains the elution sample.

The results showed that the entire protein sample loaded on the column was washed off, during the washing step. This indicated that hTP did not bind to the gel bed and probably requires further experimental investigation.

3.9.7 Crystallisation trials for native hTP and hTP mutants

3.9.7.1 Native hTP

Crystals of native hTP were previously obtained in 0.1M sodium-acetate (pH 5.2), 25% PEG 4000 and 0.1M ammonium acetate, while crystals for hTP-5IUR complex were obtained in 2.8-3.5M sodium-formate. Hence the first step was to check whether these conditions would be reproducible using a new 'batch' of hTP. The concentration of the protein sample used for crystallisation experiments was measured by BCA (PIERCE) to be ~7.7mg/ml. The condition that has previously yielded crystals of native hTP did not produce any, while that of hTP-5IUR complex produced overlapping two-dimensional plates (Figure 3.34).

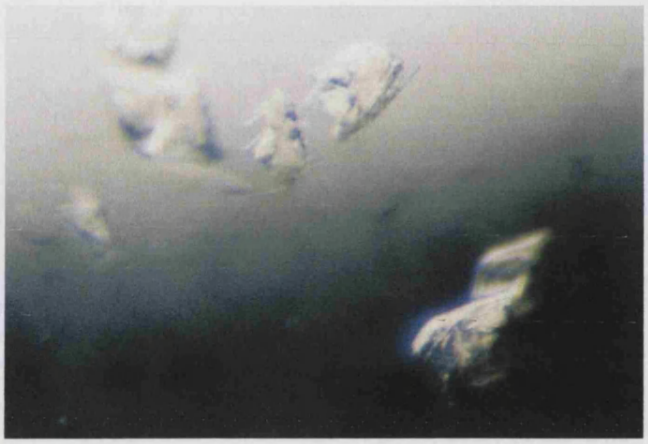


Figure 3.34: Crystals of native hTP obtained in 2.8M sodium formate. The crystals look like overlapping plates, small in size and with no clear edges.

The crystals appeared after ~70hrs incubation at 16°C. It was thus assumed that nucleation occurs quickly and in order to produce larger, single crystals, the nucleation process should be slowed-down. For that reason a series of additives and precipitants were tested in addition to screening concentrations of Na-formate (Table 3.7). All the additives and precipitants were tested separately on the initial condition (and the concentration screen) as well as in combinations in an effort to obtain single, good quality crystals. Unfortunately all conditions yielded the same type of overlapping plates (sometimes of better quality) which however could not be used for X-ray diffraction data collection.

Another commonly used method for controlling nucleation is the crystallisation under oil. For this technique 200µl of 1:1 paraffin : silicon oil (Qiagen) were added to the reservoir solution (Table 3.6. The ratio of protein: reservoir solution was 1:1 and the drops were incubated at 16°C. Unfortunately this technique did not yield any crystals.

To improve crystal quality, seeding experiments were performed. Spontaneous nucleation (without seeds present) requires a high degree of supersaturation and can sometimes lead to aggregation. When seeds are provided, more molecules can associate with them and that can lead to a 'proper', single crystal. Two types of seeding experiments were carried

out for optimising this condition: streak-seeding and micro-seeding. Streak-seeding (during which a whisker is used to touch an already formed crystal and then seeds are transferred to pre-equilibrated drops) was carried out on drops which had been pre-equilibrated with protein sample for 1hr and for 24 hrs. The ratio of protein to reservoir solution used for equilibration of the drops was from 0.5:1, 1:1, 1:2, 1:3 to 1:4 and all drops were incubated at 16°C. Micro-seeding (during which a crystal is dissolved in mother liquor and then used to provide seeds for a new drop) was carried out on drops which were also pre-equilibrated with protein sample, at the same ratios as those used for streak-seeding. Both techniques were not successful: streak-seeding produced the same type of crystals as those initially observed (overlapping, two-dimensional plates) while micro-seeding did not yield any crystals.

Table 3.6: Crystallisation trials for native hTP

Condition	**Additives and precipitants**
2.8-3.5M Na-Formate	Ethylene glycol, Glycerol, MPD, Jeffamine 600, Spermine, PEG 400, 4K, 8K, 10K, Imidazole, DMSO, $(NH_4)_2SO_4$, Na-citrate, Na-acetate, $LiCl_2$, $MgCl_2$, $CdCl_2$, $CaCl_2$
	Techniques
	Crystallisation under oil
	1:1 paraffin : silicon oil
	Streak seeding
	Micro-seeding

During these optimisation trials, structure screens I and II (Hampton Research) were also set up in order to identify new possible 'hit' conditions. Two conditions that produced crystals were: 1) 2% Dioxane, 01.M Bicine pH 9.0 and 10% PEG 20K and 2) 0.4M ammonium dihydrogen phosphate. The first condition produced small but good quality rectangular crystals. These crystals were very fragile and not reproducible. The second condition produced very thin needle crystals (Figure 3.35).

Figure 3.35: Crystals of native hTP grown in 0.4M ammonium dihydrogen phosphate and incubated at 16°C.

This condition was screened for concentrations of 0.2M to 0.5M ammonium dihydrogen phosphate and a pH range of 4.0 to 6.0. The most optimum concentration was found to be 0.4M while the most optimum pH was found to be 4.0. Further optimisation of this condition included the use of additives, precipitants and alcohols. All crystals produced resembled those obtained by the initial condition: very thin, needle crystals (Figure 3.36).

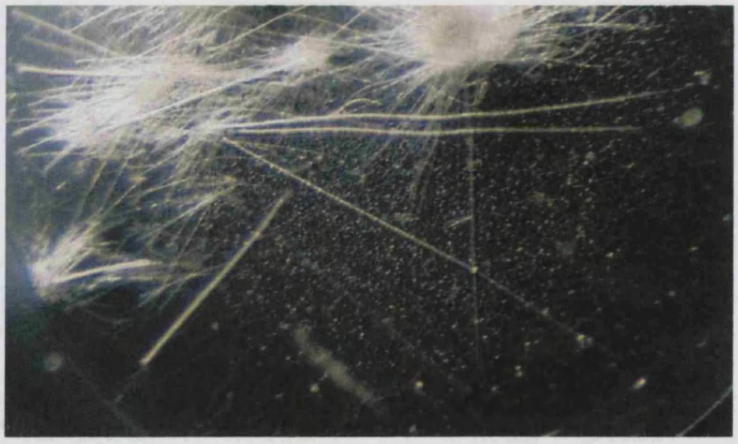


Figure 3.36: Needle crystals of native hTP obtained in 0.4M ammonium dihydrogen phosphate, pH 4.0, incubated at 16°C.

The exposure of these needle crystals to the X-ray beam revealed that they diffracted very poorly, to ~9Å. To improve crystal quality, experiments such as cross-linking (using glutaraldehyde) and dehydration were carried out. The use of glutaraldehyde did not

destroy the crystals, which even after three hours incubation maintained their original shape and size. The cross-linked crystals diffracted better, to ~6Å which however required further optimisation. Dehydration experiments did not significantly affect the crystal morphology or the diffraction pattern obtained. Experiments of the same condition were also set up for incubation at 4°C, without any success.

The use of ammonium dihydrogen phosphate in combination with sodium formate (200mM and 300mM) initially appeared to cause phase separation. The drop morphology changed however after incubation at 16°C for approximately nine months and rock – like crystals appeared in the drop (Figure 3.37).

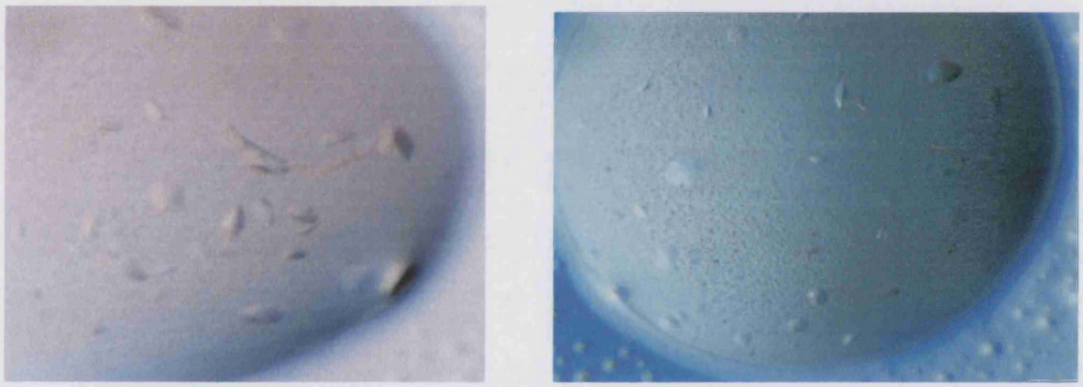


Figure 3.37: Crystals of native hTP grown in ammonium dihydrogen phosphate and sodium formate, which appeared after nine months incubation at 16°C.

These crystals however were grown under a film, and when tested under the X-ray beam proved to be very fragile and did not diffract.

3.9.7.2 Crystallisation trials for hTP mutants

Of all hTP mutants expressed and purified some presented greater 'structural' interest than others. For this reason crystallisation trials for hTP mutants started off from specific mutations rather than all of them. In order to cover all residues mutated, one 'representative' was chosen. Mutants K115E, H116F and H116K, Y199A and Y199F

were the mutations taken to crystallisation trials. All mutants were concentrated to ~8mg/ml and the drops were all set at a ratio of 1:1, protein to reservoir solution.

The first conditions to be tested for all mutants were those which provided crystals for native hTP. hTP mutants however did not crystallise in any of these conditions. In order to obtain initial hits which would be further optimised, structure screens I and II (Hampton Research) were carried out. Mutants H116F, H116K and Y199F formed clusters of needle crystals (Figure 3.38) in 0.1M $CdCl_2$, 0.1M Na-acetate at pH 4.6 and 30% PEG 400.

Figure 3.38: Cluster of needle crystals for mutant H116F in 0.1M $CdCl_2$ 0.1M Na-acetate at pH 4.6 and 30% PEG 400. Drops were incubated at 16°C.

Similar clusters of needles were obtained for H116K and Y116F.

A screen of chloride salts and PEG was set up to optimise the initial condition (Table 3.7). This screen yielded crystals of similar quality to those initially obtained, while some of the conditions did not produce any crystals.

Table 3.7: Crystallisation trials for mutants H116F, H116K and Y116F	
Condition	**Additives and precipitants**
0.1M $CdCl_2$ 0.1M Na-acetate pH 4.6 30% PEG 400	$CdCl_2$, CsCl, LiCl, $MgCl_2$, $MnCl_2$, KCl, $ZnCl_2$, $FeCl_3$, $CoCl_2$, $NiCl_2$, $CuCl_2$, $CaCl_2$, PEG 2K, MPD, Glycerol

Mutants K115E and Y199A formed crystals under completely different conditions. Initial hits (crystalloids) were obtained in 0.1M Sodium acetate, 0.1M Tris HCl pH 8.5 and 30% PEG 4K (structure screen II, Hampton Research). This condition was further optimised by screening around the concentration of PEG and sodium acetate. Additives and precipitants were also tested (Table 3.8). The drops were incubated at 16°C and the ratio of protein to reservoir was 1:1. Additives were separately tested on the initial hit as well as in combinations, in order to identify a condition which would yield good quality diffracting crystals of the mutants. A different buffer around the same pH range was also tested in order to check whether it was the pH or the buffer composition which was essential for crystal formation.

Table 3.8: Crystallisation trials for mutants K115E and Y116A	
Condition	**Additives and precipitants**
0.1M Sodium acetate 0.1M Tris HCl pH 8.5 30% PEG 4K	$CdCl_2$, CsCl, LiCl, $MgCl_2$, $MnCl_2$, KCl, $ZnCl_2$, $FeCl_3$, $CoCl_2$, PEG 400, PEG 2K, MPD, Glycerol, Jeffamine 600, Ethylene glycol, Dioxane, $(NH_4)_2SO_4$

	Buffer
	0.1M Na-HEPES pH 7.5

Optimisation of the original condition yielded either amorphous spherulites (Y199A), or clusters of needles of very poor quality (K115E; Figure 3.39) in 0.1M Sodium acetate,

0.1M Tris HCl pH 8.5, 40% PEG 4K and 0.1M Sodium acetate, 0.1M Tris HCl pH 8.5, 30% PEG 4K and 100mM $(NH_4)_2SO_4$.

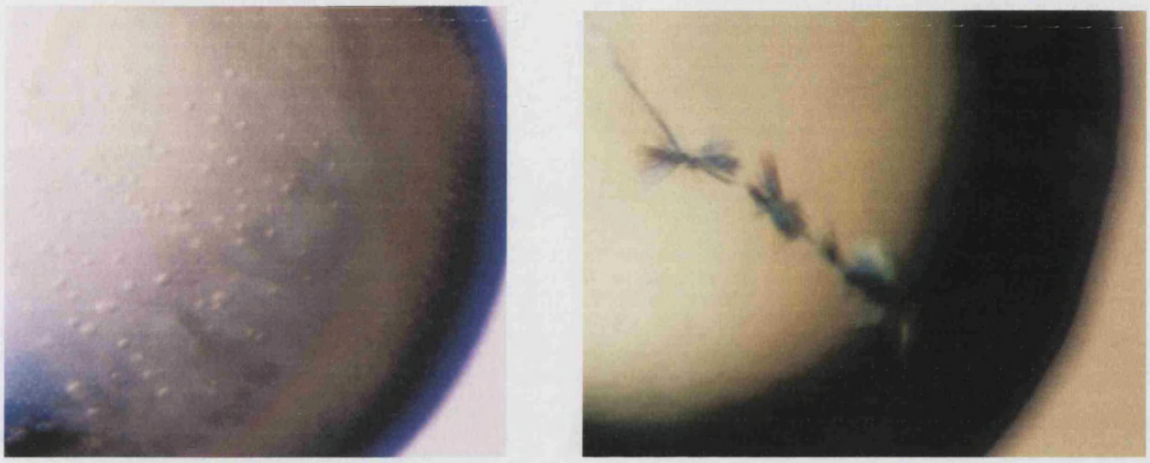


Figure 3.39: Amorphous spherulites and clusters of needles around a cloth fiber were initially obtained for mutants K115E and Y199A.

Further optimisation of this condition and substitution of Tris HCl pH 8.5 by Na-HEPES at pH 7.5 resulted in crystal formation. Crystals were observed for both K115E and Y199A in 0.1M Na-HEPES pH 7.5, 0.2M $CdCl_2$, 28% PEG 400 and 30% MPD (Figure 3.40).

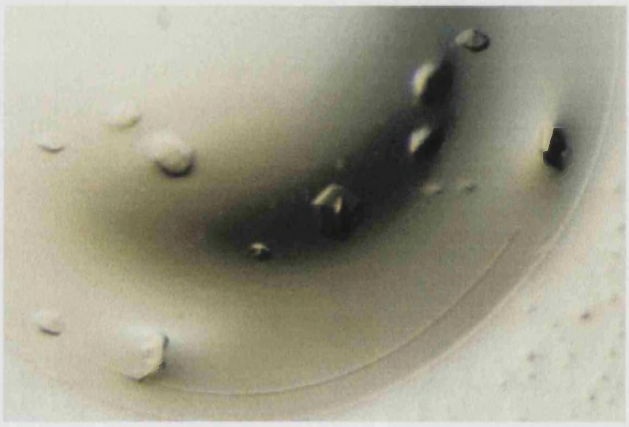


Figure 3.40: Crystals obtained for mutants K115E and Y199A in 0.1M Na-HEPES pH 7.5, 0.2M $CdCl_2$, 28% PEG 400 and 30% MPD, after incubation at 16°C.

The crystals were exposed to the X-ray beam but no diffraction was obtained. All hTP crystals, native and mutants, proved to be extremely fragile and were destroyed very quickly when exposed to the X-ray beam.

In an effort to further optimise the conditions obtained for all hTP mutants, an OptiSalt screen (Qiagen) was performed. Crystals obtained by two conditions of the screen (0.1M Na-acetate pH 4.6, 0.6M Na-fluoride and 0.1M Tris HCl pH 8.5, 0.6M Na-fluoride) turned out to be salt crystals.

Protein for mutants R202E, R202S, S217G and I214A was produced much later than that of the previous mutations on K115, H116 and Y199. For this reason, crystallisation experiments on R202, S217 and I214 mutants are not described.

3.9.8 Enzymatic activity assays and correlation with hTP mutant models

The phosphorolytic activity of hTP was assessed by enzymatic activity assays, as described above (Chapter III, *3.8.11*). Analysis of the results obtained from the assays indicated that native hTP (His-tag on the N-terminal domain) was active, with a k_{cat} of $8.2s^{-1}$ (Table 3.9), comparable to that reported previously by Finnis, where $k_{cat} = 9.4s^{-1}$ for recombinant native hTP (no His-tag on the N-terminal domain; Finnis *et al.*, 1993). The profile of the phosphorolytic reaction catalysed by native hTP is given by the plot of velocity V against substrate concentration [S] (thymidine); Figure 3.41.

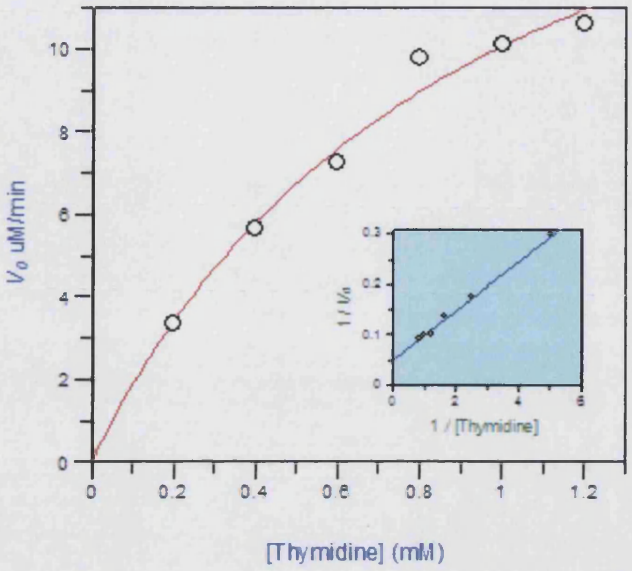


Figure 3.41: Phosphorolytic reaction catalysed by native hTP. The plot of velocity V against substrate concentration [S] here illustrates the profile of the reaction. Shown in light blue is the double reciprocal plot (Lineweaver-Burk plot) of $1/V$ against 1/[S].

Of all hTP mutants (total of eleven point mutations; K115E, K115A, H116K, H116F, Y199A, Y199L, Y199F, R202E, R202S, S217G and I214A) only four retained some of the enzymatic activity of hTP. K_m, V_{max} and k_{cat} for hTP and for each of the mutants are given in Table 3.9. Two hTP mutants, K115E and R202S, previously reported by Miyadera and co-workers (Miyadera *et al.*, 1995) were repeated to confirm the previous results and to provide a standard for the rest of the mutations carried out.

Table 3.9 : Enzymatic activity assay results for native hTP and hTP mutants

	K*m* (mM) (Std. Error)	V*max* (μM/min) (Std. Error)	k*cat* (s^{-1})
Native hTP	0.35 (0.01)	19.7 (2)	8.2
K115E	-	-	-
K115A	0.42 (0.04)	16.4 (0.9)	6.8
H116F	-	-	-
H116K	-	-	-
Y199A	-	-	-
Y199L	3.14 (1.0)	10.6 (2)	4.4
Y199F	2.1 (0.6)	8.4 (1)	3.5
R202E	-	-	-
R202S	-	-	-
I214A	0.74 (0.03)	12.53 (0.3)	5.2
S217G	-	-	-

No product formation (thymine) could be detected for mutant K115E, signifying that the mutation had lead to complete loss of enzymatic activity. This result is in agreement with previous work by Miyadera and co-workers (Miyadera *et al.*, 1995), where mutant K115E had no enzymatic and angiogenic activity. On the other hand, mutant K115A reduced the enzymatic activity of hTP by about 1.2 fold. V_{max} for K115A (16.4μM/min; Table 3.9; profile for phosphorolytic reaction of K115A is shown in Figure 3.42) is slightly lower than that of the native enzyme, indicating that this mutation has a mild effect on the enzymatic activity of hTP. Interestingly, the van der Waals contacts for K115E appear to be more than those for K115A. Lys115 in native hTP is involved in 21

van der Waals contacts, which are reduced to twelve in K115A, but increased to twenty nine in K115E (Table 3.10). Although there is an increase in the number of interactions for K115E as opposed to Lys115 of native hTP, it is very likely that introduction of negative charge at this position does not facilitate the binding of phosphate due to charge repulsion, thus rendering the enzyme inactive. Mutant K115A loses the positive charge of Lys but introduces a hydrophobic residue which appears to retain some of the enzymatic activity.

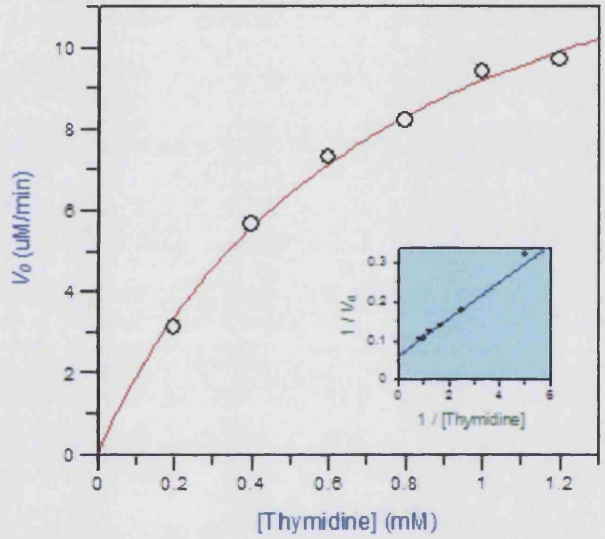


Figure 3.42: Phosphorolytic reaction catalysed by hTP mutant K115A. The plot of velocity *V* against substrate concentration [S] illustrates the profile of the reaction. Shown in light blue is the double reciprocal plot (Lineweaver-Burk plot) of $1/V$ against 1/[S]. It is evident that K115A reduces the velocity of the reaction, without however completely abolishing it.

Table 3.10: Hydrogen bond and van der Waals interactions for native hTP and hTP mutants.

hTP-5IUR complex	No. of van der Waals contacts	No. of hydrogen bonds
K115	**21**	**-**
H116	**30**	**2**
Y199	**38**	**3**
R202	**29**	**3**
I214	**38**	**2**
S217	**28**	**3**
hTP mutants		
K115A	**12**	**-**
K115E	**29**	**-**
H116F	**39**	**1**
H116K	**29**	**1**
Y199A	**19**	**2**
Y199F	**38**	**2**
Y199L	**32**	**2**
R202E	**22**	**-**
R202S	**15**	**1**
I214A	**14**	**2**
S217G	**15**	**2**

Shown in this table are the van der Waals contacts and the hydrogen bonds for hTP-5IUR and each of the mutations. van der Waals interactions have been calculated for each individual residue both in hTP-5IUR and mutants. Also shown here, are the hydrogen bond interactions for each individual residue.

The significance of His116 has been discussed previously, on *E.coli* thymidine phosphorylase (Rick *et al.*, 1999) and in hTP (Norman *et al.*, 2004), although mutational work on this residue has not been reported so far. Enzymatic activity assays carried out on two His116 mutants (H116F and H116K) indicate that both mutations have a severe effect on hTP, as they completely abolish the enzymatic activity (Table 3.9). Analysis of intramolecular interactions in these mutants showed a decrease in the number of hydrogen bonds as opposed to native hTP; from 2 (native) to 1 (H116K and H116F). van der Waals interactions increased in mutant H116F, while in H116K these interactions were reduced slightly (Table 3.10). Since both His116 mutations completely abolish the enzymatic activity of hTP, it is possible that both the basic charge and the ring structure of histidine at this position are an absolute requirement and that His116 is an absolutely crucial residue for the catalytic activity of hTP. Structural studies of the two His116 mutants would possibly reveal whether the binding of substrate in the active site is hindered when His116 is mutated.

Tyr199, whose role is yet unclear, appears to affect the catalytic activity of hTP. Mutations on Tyr199 provided some interesting information on its role and significance for the catalytic activity of hTP. Among the three tyrosine mutations, Y199L and Y199F reduced the catalytic activity of hTP, while Y199A completely abolished it (Table 3.9). Mutant Y199L catalyses the reaction at a much slower rate than native hTP ($V_{max\ Y199L}$ 10.6μM/min; Table 3.9; Figure 3.43) while its catalytic efficacy is almost half of that observed for native hTP ($k_{cat} = 4.4s^{-1}$; Table 3.9).

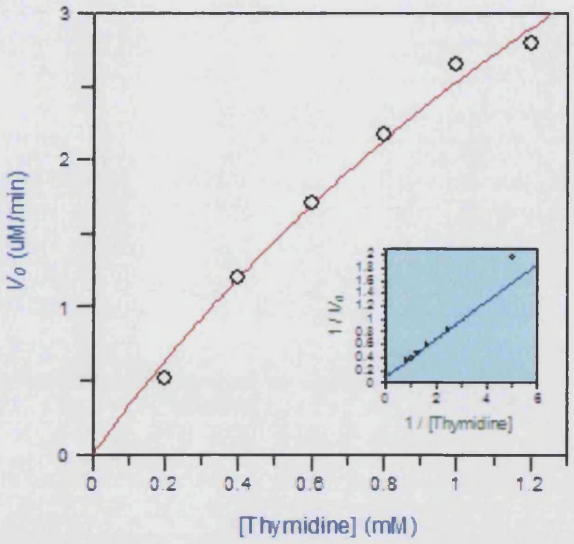


Figure 3.43: Phosphorolytic reaction catalysed by hTP mutant Y199L. The plot of velocity V against substrate concentration [S] illustrates the profile of the reaction. Shown in light blue is the double reciprocal plot (Lineweaver-Burk plot) of $1/V$ against 1/[S]. Mutant Y199L was shown to reduce the rate of the phosphorolytic reaction and the catalytic efficacy (Table 3.9).

Mutant Y199F reduced the rate of the reaction even more than mutant Y199L ($V_{max\ Y199F}$ 8.4μM/min; Table 3.9; Figure 3.44) as well as the catalytic efficacy (reduced to $3.5s^{-1}$ for Y199F). The difference between the results obtained by the two Tyr199 mutants are however within error limits, as seen from Table 3.9.

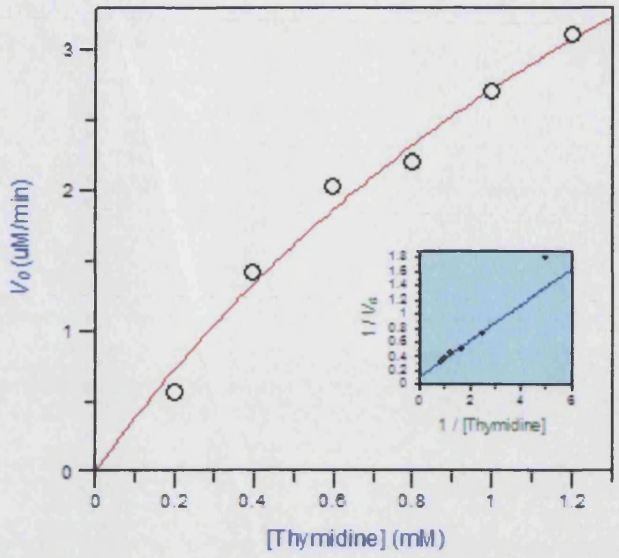


Figure 3.44: Phosphorolytic reaction catalysed by hTP mutant Y199F. The plot of velocity *V* against substrate concentration [S] illustrates the profile of the reaction. Shown in light blue is the double reciprocal plot (Lineweaver-Burk plot) of $1/V$ against 1/[S]. Mutant Y199F significantly reduced the rate of the phosphorolytic reaction and the catalytic efficacy (Table 3.9).

Analysis of the models generated for Y199F, Y199L and Y199A shows a reduction in the number of putative interactions (Table 3.10). More specifically, hydrogen bonds are reduced for all mutants, while van der Waals interactions are almost halved in mutant Y199A, reduced in mutant Y199L and retained in mutant Y199F, the latter being a rather conservative mutation as it retains the ring structure of the residue albeit the loss of the hydroxyl group. Examining the environment surrounding Tyr199 in the hTP-5IUR complex shows that the ring structure is essential for a number of intramolecular contacts in order to maintain the integrity of the active site. van der Waals interactions of Tyr residue include those with Leu148 which is a part of the hydrophobic pocket holding the iodine atom of 5IUR in place (or chlorine atom of TPI in 1UOU; Norman *et al.*, 2004). Leu148 in turn interacts with Val208, also part of the same hydrophobic pocket. Mutation of Tyr199 to Phe retains these interactions, thereby retaining the enzymatic activity of hTP. It is very likely that both size and hydrophobicity of the residue at this position is

essential for maintaining the structural integrity of the active site. Any change otherwise disfavours the normal function of hTP. It is thus evident that Tyr199 is indeed involved in the enzymatic activity of hTP.

Another mutation which completely abolished the enzymatic activity of hTP was the one targeting residue Arg202. Arg202 is involved in the hydrogen bonding network that stabilises 5IUR in the active site cleft, as seen in the hTP-5IUR complex structure (Figure 3.18). Site-directed mutagenesis carried out previously on Arg202, (R202S by Miyadera *et al.*, 1995) resulted in abolishing the enzymatic and angiogenic activity of hTP. Our findings are in agreement with previous work; both Arg202 mutants (R202E and R202S) completely abolish the enzymatic activity of hTP (Table 3.9). Models of the two Arg202 mutants indicate a significant reduction both in hydrogen bonds and van der Waals interactions. In native hTP, Arg202 is involved in three hydrogen bonds and a total of twenty nine van der Waals interactions. Mutant R202E shows a reduction in van der Waals contacts to twenty two, while no hydrogen bonds could be detected for the mutated residue. Mutant R202S shows an even more drastic reduction in the number of van der Waals interactions, reduced to only fifteen (Table 3.10). In contrast to R202E, R202S retains one of the three hydrogen bonds, observed in the native hTP (Table 3.10).

Ser217 is also part of the active site cleft and has been shown to be involved in stabilising the molecules bound in the active site of the enzyme *via* hydrogen bonds (Figure 3.18; hTP-5IUR complex structure; Norman *et al.*, 2004). No mutational work has so far been reported for Ser217 and hence no information was available on its significance for the enzymatic activity. In the present study, Ser217 was mutated to glycine and the mutant was assayed for detection of enzymatic activity. No thymine production could however be detected, indicating that the mutation had abolished the enzymatic activity (Table 3.9). Analysis of the intramolecular contacts of S217G indicated a significant reduction in the number of van der Waals interactions and a loss of one hydrogen bond, as opposed to native hTP (Table 3.10). This hydrogen bond, involving the hydroxyl group of Ser contributes to the stabilisation of the bound molecule in the active site cleft, and appears to be essential for maintaining the enzymatic activity of hTP.

Site-directed mutagenesis on Leu148 (Miyadera *et al.*, 1995), part of the hydrophobic pocket of hTP (Figure 3.19), indicated its significance for the enzymatic activity of hTP, as the mutation completely abolished both the phosphorolytic and angiogenic activity of the enzyme. Mutation of Ile214, another member of the hydrophobic pocket, showed reduction but not loss of the enzymatic activity. More specifically, mutant I214A reduced the rate of the phosphorolytic reaction, ($V_{max\ I214A}$ 12.5μM/min; Table 3.9), as well as the catalytic efficacy of hTP ($k_{cat} = 5.2s^{-1}$; Table 3.9). Analysis of the network of interactions for native Ile214 and I214A indicates a significant reduction in the number of van der Waals contacts while hydrogen bond interactions are retained (Table 3.10). Although the reduction in intramolecular interactions is significant, mutant I214A does not abolish the enzymatic activity of hTP, probably due to the retaining of the hydrophobic character of Ile by Ala. The reduction, on the other hand, of the catalytic efficacy is most likely due to the loss of the 'bulkier' side chain of Ile to Ala.

A summary of the results obtained by the enzymatic activity assays for native hTP and each of the mutants is given in Figure 3.45.

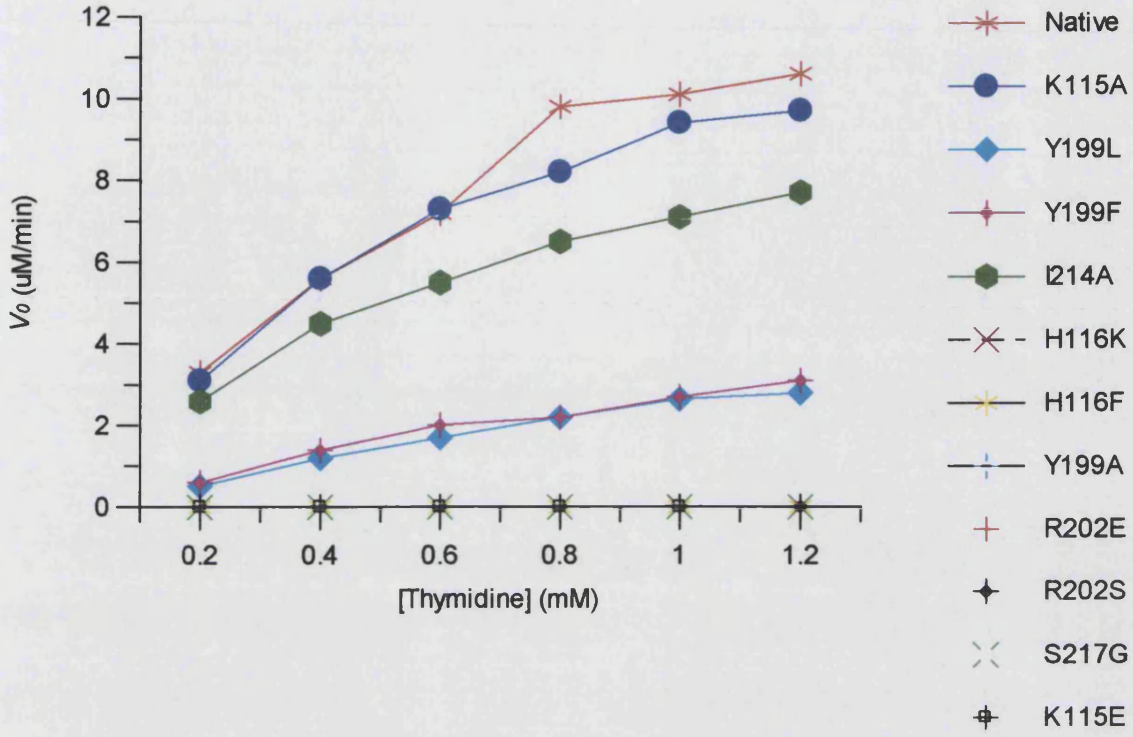


Figure 3.45: Phosphorolytic reaction catalysed by native hTP and hTP mutants. The plot of velocity *V* against substrate concentration [S] illustrates the profile of the reaction. Mutants K115A, Y199L, Y199F and I214A maintain some of the enzymatic activity of hTP, while mutants K115E, H116K, H116F, Y199A, R202E, R202S and S217G abolish the catalytic activity of the enzyme.

The catalytic efficacy of native hTP and hTP mutants, as determined by the enzymatic activity assay, were used to plot a graph (Figure 3.46), which collectively shows the results obtained.

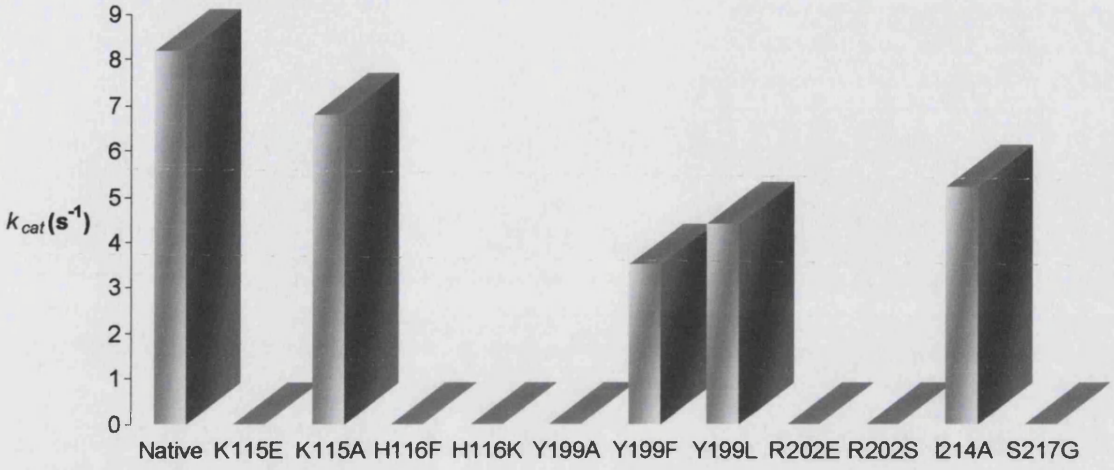


Figure 3.46: Catalytic efficacy of native hTP and hTP mutants. Mutations K115A, I214A, Y199F and Y199L reduce the enzymatic activity of hTP, with the first two having less of an effect on the catalytic potency of the enzyme. Mutants K115E, H116K, H116F, Y199A, R202E, R202S and S217G completely abolish the catalytic activity of the enzyme, as also seen from Figure 3.41.

The results of the enzymatic activity assays illustrate the significance and indispensability of certain residues to the catalytic activity of hTP. It is evident that His116, Arg202 and Ser217 are absolutely essential for the enzymatic activity. It is important to note that hTP was rendered inactive when mutations on residues which are involved in hydrogen bonding interactions with the molecules bound in the active site of the enzyme were introduced.

Lys115 affects the enzymatic activity mostly due to its interactions with phosphate, essential for the phosphorolytic reaction. Mutagenesis on Tyr199 affects but not always abolishes the enzymatic activity of hTP. It is likely that certain features surrounding the active site cleft can be conserved, by conservative mutations, which affect without abolishing the activity of the enzyme. It is thus possible to suggest that it is not only the active site residues which are essential for the binding of the substrate, but the

hydrophobic pocket and the phosphate-binding site residues, which collectively retain the catalytic activity of hTP.

3.9.9 Significance of phosphate binding in hTP activity

As one of the substrates for the phosphorolytic reaction catalysed by hTP, phosphate is required for the catalysis that yields thymine and 2dDR-1P. The role of phosphate in the reaction catalysed by hTP has been previously discussed by Birck and Schramm (Birck and Schramm, 2004) where enzymatic activity assays which omitted phosphate, resulted in a very low rate of reaction, resembling that of an inactive enzyme. The detection of some activity was probably the result of the hTP 'acting' as a hydrolase, as which however the enzyme does not seem to be very efficient. The role of phosphate became however a bit unclear, as in a recent structure where hTP was crystallised with thymidine, the product thymine was present in the active site of the molecule as seen in the crystal structure, without phosphate being included in the crystallisation conditions (PDB code 2J0F; El Omari *et al.*, 2006). This raises questions about the necessity of phosphate for the catalysis of the reaction. One crystal structure where phosphate was present in the phosphate binding site was that of BsPyNP (Pugmire and Ealick, 1998). Crystal structure of BsPyNP (PDB ID 1BRW; Pugmire and Ealick, 1998) was aligned against those of hTP-TDR (PDB code: 2J0F; El Omari *et al.*, 2006) and hTP-5IUR. This alignment reveals a water molecule at the phosphate binding site, in both the hTP structures. It is possible that this water molecule mediated the nucleophilic attack on the sugar moiety of thymidine just as phosphate would have if it were present, in the case of hTP-TDR. The sole possible source of phosphate in hTP-TDR complex which could mediate the catalysis, would be from the culture media, assuming phosphate was present and was not removed in subsequent purification steps. In such an instance, phosphate could mediate the reaction and subsequently leave the phosphate-binding site.

In order to understand the necessity of phosphate for catalysis an enzymatic activity assay was carried out for native hTP, this time omitting the phosphate source, KH_2PO_4. If phosphate is an absolute requirement for the reaction to occur then no product (thymine) should be detected. On the other hand, if water can mediate the reaction and hTP is a

potent hydrolase, then thymine would still be detected, even in the absence of phosphate. This would also be the case if phosphate is already present from the expression media.

Results from the enzymatic activity assay confirmed the previous work which suggested hTP is not a potent hydrolase. More specifically, very low enzymatic activity was detected when the phosphate source was removed, resembling an almost inactive enzyme (K_m, V_{max} and k_{cat} values were too low to be determined). This result suggests that the hydrolytic activity of hTP is very low especially when compared to the phosphorolytic activity of the enzyme. Since phosphate is required for the reaction to occur it is not likely that water molecules seen in the phosphate binding site would mediate an effective hydrolytic reaction. It is thus evident that phosphate is an absolute requirement for the reaction catalysed by hTP and upon the removal of phosphate, the enzyme becomes inactive.

3.9.10 Is 5IUR a potent hTP inhibitor?

In the present structure of hTP-5IUR complex, the small size of the inhibitor as well as its similarity to 5FUR (currently used chemotherapeutic agent) make it an attractive target for drug design. Kinetic studies carried out by Razzell and Casshyap (1964) have shown a 61% inhibition of thymidine arsenolysis when 0.64mM of 5IUR was used (Razzell and Casshyap, 1964). In order to determine the type of inhibition and the potency of 5IUR in inhibiting hTP, enzymatic activity assays, including 5IUR, were carried out. The substrate concentrations used for this assay were (0.2mM, 0.6mM and 1.2mM), while inhibitor concentrations used were 0.005, 0.01, 0.05, 0.1, 0.15 and 0.2mM. The inhibition of phosphorolysis was measured by the reduction in the reaction rate. To determine the type of inhibition and the K_i for 5IUR Dixon plots were used (Appendix II; Dixon, 1953). Results for 1/V and [S]/V were plotted against the inhibitor concentration, for the different substrate concentrations (Figure 3.47, Figure 3.48).

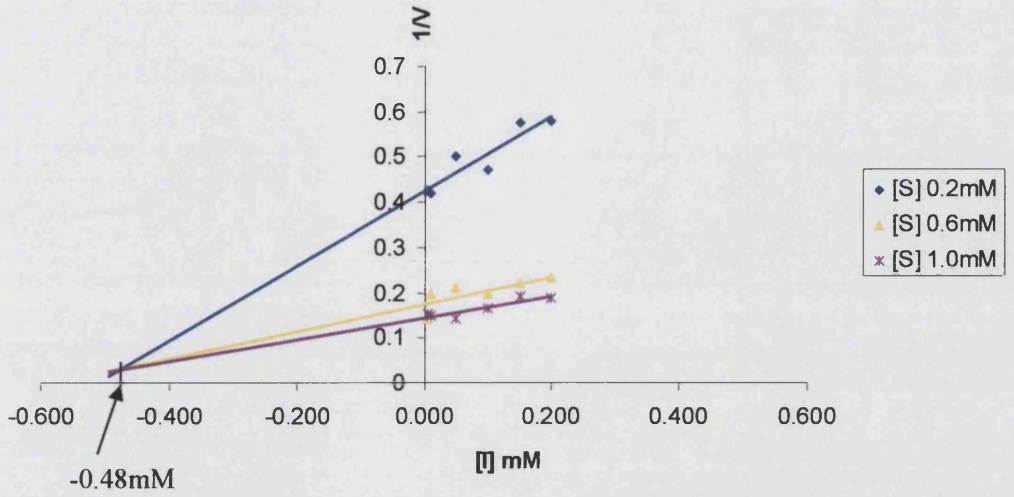


Figure 3.47: Plot for 1/V against the inhibitor concentration, [I]; the concentration of 5IUR. The three substrate concentrations used for these measurements are 0.2mM, 0.6mM and 1.0mM. The three substrate concentrations 'meet' just above the x-axis, which is indicative of competitive inhibition.

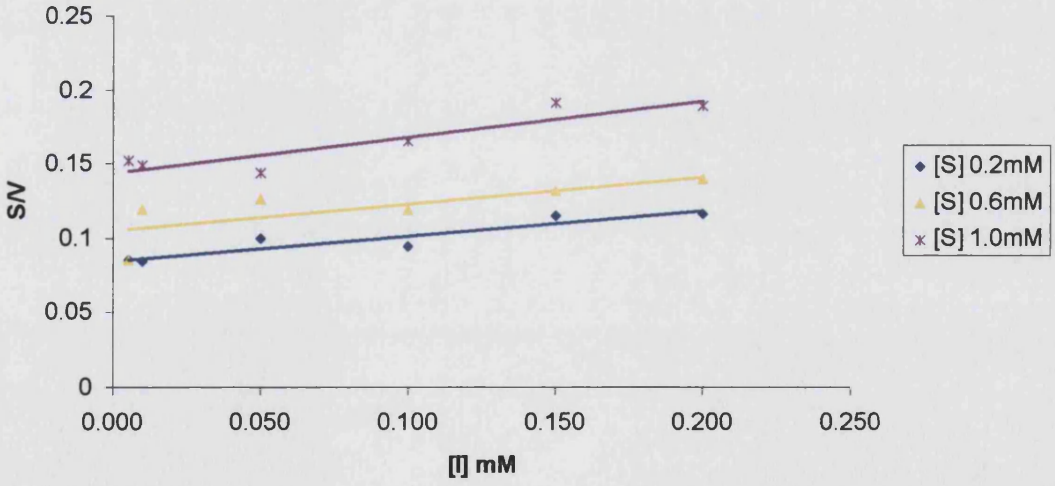


Figure 3.48: Plot of [S]/V against the inhibitor concentration, [I]. The measurements at three substrate concentrations produce parallel lines, which is indicative of competitive inhibition.

According to the published method of Butterworth (Butterworth, 1972) where Dixon plots are presented as way to study enzyme inhibition, the plots of 1/V (Figure 3.47) and [S]/V (Figure 3.48) were generated. In competitive inhibition, the substrate concentrations in a plot of 1/V against [I] should intersect above the x-axis, for the different substrate concentrations, while in the plot of [S]/V against [I] substrate concentrations should be parallel. The results from the inhibition assay for 5IUR are in agreement with this type of inhibition and it is thus possible to say that 5IUR is a competitive hTP inhibitor. Due to its similarity to the product of the phosphorolytic reaction catalysed by hTP, 5IUR would be expected to 'compete' with the substrate for the active site of the enzyme.

Dixon plots are an efficient way for determining the inhibition efficiency, K_i. In competitive inhibition K_i can be determined from the plot of 1/V against [I]. According to the results obtained for 1/V against [I], K_i is 0.48mM. This K_i is 'low' for a competitive inhibitor, but it is in agreement with previous results as reported by Razzell and Casshyap (Razzell and Casshyap, 1964), where only 61% inhibition of hTP arsenolysis was achieved when 0.64mM of 5IUR were used. Hence, although a competitive inhibitor, 5IUR is not a very efficient hTP inhibitor. It is possible however that modification on 5IUR would increase its efficacy, which along with the lack of a toxicity group (such as the fluorine atom in 5FU) would probably make it an interesting drug target.

3.9.11 Conclusion and future work

Thymidine phosphorylase is an enzyme responsible for the reversible phosphorolysis of thymidine to thymine and 2-deoxy-D-ribose-1-phosphate. The catalytic activity of TP, essential for providing pyrimidine nucleotides for DNA repair and replication, has been shown to be indispensable for the angiogenic activity of the enzyme. The role of hTP in angiogenesis (physiological and pathological) has attracted a lot of interest for the enzyme. In physiological angiogenesis, its role as the 'supplier' of pyrimidine nucleotides is essential for the salvage pathway and for maintaining the nucleic acid homeostasis. Several pathological conditions have also been linked to hTP: detection of the enzyme in primary tumours is a risk factor as it has been associated with tumour invasiveness and

metastasis. hTP has also been found to overexpress in conditions such as rheumatoid arthritis, ulcers, irritable bowel disease and psoriatic lesions, making it thus a primary target for design and development of new drugs. As the angiogenic activity of hTP is dependent on its enzymatic activity, it is essential to target the enzymatic activity in order to inhibit its role in angiogenesis. Inhibitors of hTP can act as anti-angiogenic agents, significantly reducing tumour growth by disrupting the thymidine salvage pathway, angiogenesis or metastasis (Esteban-Gamboa *et al.*, 2000; Kita *et al.*, 2001). Several inhibitors have been designed for hTP and have mostly been tested in kinetic studies. The most potent inhibitor of hTP reported to date is TPI (IC_{50} value of 35nM), proposed to suppress tumour growth by increasing the apoptotic rate. Currently used as a chemotherapeutic agent target the activity of hTP is 5FU, administered as an oral prodrug, doxifluridine, which however is a rate-limiting drug as toxicity is accumulated not only in tumour areas but in the gastrointestinal tract as well. A new drug, currently in clinical trials, capecitabine, was designed so as for the prodrug to pass through the gastrointestinal tract at its intact form and deliver the converted 5FU to cancer cells only (Miwa *et al.*, 1998).

Structural information has greatly enhanced the understanding of the structure and function of the members of the PyNP superfamily. The first structures to be reported were those of EcTP (Walter *et al.*, 1990) and BsPyNP (Pugmire and Ealick, 1998), followed by their human homologue. Information gained from structural and computational studies on EcTP revealed two states of the enzyme: a bound-closed conformation and an unbound-open conformation. Analysis of the reported structures of TP from different species, indicates conservation in several characteristics, but also raises questions about some of the proposed mechanistic details employed for the catalytic activity of TP.

Two crystal structures of hTP have been reported to date: a complex of hTP-TPI, a small molecule inhibitor (PDB ID 1UOU; Norman *et al.*, 2004) and a complex of hTP-TDR, the product of the phosphorolytic reaction catalysed by hTP, thymine (PDB ID 2J0F; El Omari *et al.*, 2006). Both structures are complexes and provide information solely about

one 'state' of the enzyme, in which it is in complex with another molecule. To understand the mechanistic details that drive the catalytic activity of hTP, more information, structural (e.g. bound and unbound forms of the enzyme) and functional (significance of active site residues) are required. The two structures of hTP, native and in complex with 5IUR, reported here, provide new insights and raise new questions about the structural and functional understanding of the enzyme.

The crystal structure of hTP-5IUR and the analysis of the network of interactions holding the small molecule inhibitor in the active site cleft, showed similarities with previously reported hTP complexes (Normal *et al.*, 2004; El Omari *et al.*, 2006). The active site residues previously identified were also detected to form the same network of interactions with 5IUR (Figure 3.18). This was however expected, as several residues that form the active site cleft are conserved among species, as seen from structures of EcTP (Walter *et al.*, 1990) and BsPyNP (Pugmire and Ealick, 1998). It is however possible that other residues, are also involved in stabilising the substrate thymidine in the active site cleft of the enzyme, which however have not been described to date as all molecules in complex with hTP reported so far are smaller in size.

Alignment of bound (hTP-5IUR) and unbound (native hTP) hTP structures, indicated a lack of 'domain movement', previously proposed for members of the PyNP superfamily. Domain movement was proposed to be essential in bringing the substrates closer together, for catalysis to occur. Based on our findings however, domain movement was not detected for the enzyme, which might indicate that it does not actually occur in the case of hTP (Figure 3.21). On the other hand, it is possible that the 'closed' state of the enzyme, observed in both bound and unbound hTP structures reported here, is favoured by crystal packing, and binding of phosphate or substrates is not required for the enzyme to acquire this particular conformation. It would not thus be accurate to state with confidence whether domain movement occurs in hTP or whether it is essential for catalysis, as structural information for possible intermediate stages have not been reported to date. Identification of intermediate stages, both functional and structural, could be used for drug design and development. It would be possible to design drugs that

target specifically the intermediate stages of the phosphorolytic reaction catalysed by hTP, so as to ensure the enzyme is not inhibited in its inactive form. This would be particularly useful as inhibition of physiological hTP activity can lead to several pathological conditions.

The potency of the inhibitors currently developed and tested against hTP has been previously discussed by Birck and Schramm (Birck and Schramm, 2004). Kinetic isotope effect measurements were used for the determination of the transition states of the reaction catalysed by hTP. Research on the transition states, using computational methods, has also been carried out by two different research groups (Cole *et al.*, 2003; Mendieta *et al.*, 2004). Computational methods yielded similar results: they identified an S_N1-type reaction for hTP, which involves the formation of a carbocation intermediate preceding the subsequent nucleophilic attack. In this type of reaction, the overall reaction rate depends on the concentration of the reactant only. A different type of reaction mechanism was however identified by the kinetic isotope effect measurements (Birck and Schramm, 2004). Experimental results indicated that hTP undergoes an S_N2-type reaction, which involves the direct displacement of a leaving group (2-deoxy-D-ribose) from the nucleophile (phosphate). This type of reaction mechanism does not involve the formation of a carbocation intermediate and is dependent on the concentration of the nucleophile as well as the concentration of the substrate. If this is in fact the reaction mechanism used in the phosphorolytic reaction which hTP catalyses, it is the first case of an S_N2-type reaction for the family of *N*-ribosyl transferases. It is evident that inhibitors which were designed to 'block' an S_N1-type reaction will not be as efficient in the inhibition of the S_N2-type reaction seen in hTP. Hence, it is possible that new small molecule inhibitors, aiming at the inhibition of an S_N2-type reaction would be more efficient than the currently designed and tested compounds. Hence, compounds designed to efficiently inhibit hTP should be substrate analogues, with reduced toxicity and specific for the reaction type employed by the enzyme. It is not yet possible to predict whether reaction type and domain movement are related. Structures of transition states of the enzyme would provide more information and would aid in the understanding of both mechanistic details and conformational changes that might occur during catalysis.

In order to understand the mechanistic details of hTP it is essential to elucidate the role and significance of the active site residues. Previous work carried out by Miyadera *et al.* indicated that residues Lys115, Arg202 and Leu148 were indispensable for the enzyme, as mutagenesis abolished both enzymatic and angiogenic activity of hTP. The role of His116 (His85 in EcTP) was previously discussed by Rick *et al.* but mutagenesis work involving this residue was not reported to date. In order thus to understand the role of active site residues and their significance for the enzymatic activity, mutagenesis was carried out on Lys115, His116, Tyr199, Arg202, Ile214 and Ser217. With the exception of Ile214, which is part of the hydrophobic pocket, the rest of the residues examined are all involved in stabilising 5IUR in the active site cleft by means of hydrogen bonds. Introduction of point mutations to hTP, expression and purification of all mutants were followed by enzymatic activity assays in order to determine their effect in the phosphorolytic efficacy of hTP.

Lys115, was previously proposed to stabilise phosphate in the phosphate binding site along with residues Ser126, Ser144 and Ser117 (Edwards, 2006). Enzymatic activity assay results indicated that Lys115 modulates the enzymatic activity of hTP *via* its interactions with phosphate, which is one of the required substrates for catalysis. It is evident that the positive charge at that particular position is the required environment for efficient phosphate binding, which in turn allows for the phosphorolysis to occur. Introduction of a negative charge changes the environment of the phosphate-binding site and yields the enzyme inactive.

The highly conserved His116 renders the enzyme completely inactive when mutated, indicating its significance for the catalytic activity of hTP. The role previously attributed to His116 in hTP (Norman *et al.,* 2004) as part of a proton shuttling mechanism could not be established with the results obtained from mutants H116F and H116K. Mutational work however introducing glutamic acid at position 116 would indicate whether His116 is in fact involved in a proton shuttling mechanism or whether it is simply involved in hydrogen bonding interactions in the active site cleft.

Mutations on Arg202 and Ser217, both of which interact directly with bound molecules at the active site of the enzyme, adversely affect the catalytic efficacy of hTP, as previously seen for His116. This is indicative of the significance of the network of interactions in which these residues are involved, for efficient binding and catalysis by hTP. In contrast to these mutations, mutagenesis on Ile214, which is part of the hydrophobic pocket of hTP, reduces the enzymatic activity without completely abolishing it. It is thus evident from the results obtained that alteration of the environment around the active site, hydrophobic pocket and phosphate-binding site affects and in some cases completely abolishes the enzymatic activity of hTP.

Tyr199, whose role was so far unclear, was identified to modulate the enzymatic activity of hTP, according to the mutation introduced at that particular position. The results obtained, indicate that the ring structure is essential at that position as mutation of Tyr to Ala completely abolished the enzymatic activity. Mutants Y199F and Y199L retained some of the enzymatic activity, which was however reduced when compared to native hTP.

The role of phosphate in the enzymatic reaction catalysed by hTP was shown to be indispensable. Enzymatic activity assay results, where phosphate was not included, resembled an inactive enzyme. The rate of reaction and catalytic efficacy were too low to be determined, indicating the significance of phosphate for the phosphorolytic reaction catalysed by hTP.

The insights gained from this functional work will provide a platform for further structural and functional studies, which would contribute towards a better understanding of the mechanistic action of hTP. Experiments such as fluorescence resonance energy transfer (FRET; Lakowicz, 1999), which allow the monitoring of protein-protein or protein-ligand interactions, would be useful in understanding the details of the reactions catalysed by hTP. Further mutagenesis work, including double and triple mutants of the active site cleft, phosphate-binding site and hydrophobic pocket, would further confirm the role of certain residues in the catalytic activity of hTP. Mutagenesis on the glycine-

rich loop, previously suggested to be involved in phosphate-binding (Walter *et al.*, 1990) would indicate whether this is in fact a conserved and essential feature of members of the PyNP superfamily. Mutants could also be subjected to circular dichroism, which allows the identification of the secondary structure characteristics. Although unlikely, it is possible that certain mutations affect the enzyme in such a way that the folding of the protein does not occur properly. Biological assays, such as cell proliferation assays, could be used to further confirm the effect of hTP mutants. More specifically, as the angiogenic activity of hTP depends on the enzymatic activity, mutants which abolish the latter would be expected to abolish the angiogenic activity as well.

The determination of crystal structures of hTP in complex with thymidine, hTP mutants and hTP in transition states would provide a wealth of information, essential for understanding the enzyme and its mechanistic details. Results from these studies, coupled with isothermal titration calorimetry (ITC; O'Brien *et al.*, 2000) would allow the design and development of more specific drugs, targeting the enzymatic activity of hTP. The implication of hTP in several pathological conditions and the drawbacks it causes in treatment of several of these cases make it a primary target for research, which aims not only at inhibiting its activity but in modulating it, especially in cases where it appears to overexpress. Its role in the nucleic acid homeostasis makes hTP an essential enzyme for the living organism which however needs to be controlled in order to maintain a physiological environment rather than a pathological condition.

Chapter – IV –

Modeling studies on Tissue Inhibitor of Metalloproteinase (TIMP) from *Drosophila melanogaster*

4.1 Introduction to the extracellular matrix

Cellular differentiation occurring in the early stages of embryogenesis, results in the formation of four major types of cells: muscle cells, epithelial cells, connective cells and neuronal cells. Each of these cell types is organised in a tissue (muscle, epithelial, connective and nervous tissue respectively), which consists of cells and the surrounding matrix, known as the extracellular matrix (ECM). ECM is mainly composed of fibrous elements, such as collagen and elastin, link proteins, such as laminin and fibronectin and glycosaminoglycans (Figure 4.1). It also contains several minerals and fluids such as blood plasma or serum respectively.

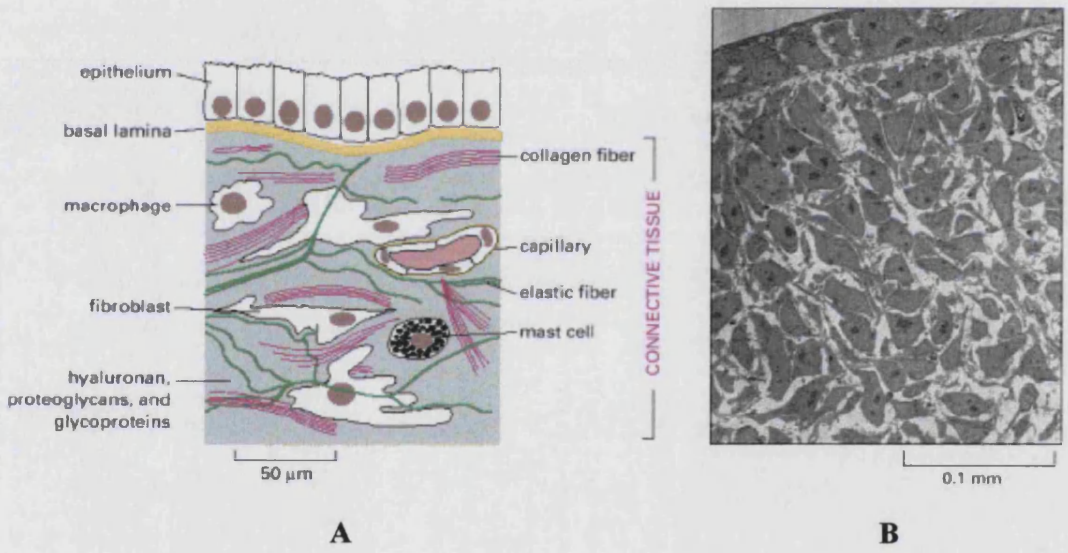


Figure 4.1: The extracellular matrix (ECM). **(A)** Illustration of connective tissue, where a variety of cells and ECM components, among which the fibroblasts, which secrete extracellular matrix. **(B)** A low-power electron micrograph of cells surrounded by extracellular matrix. The ECM supports and holds tissues and organs together providing thus an essential means for maintaining a balanced environment for cell-cell and cell-matrix interactions.(Figures adapted from: Alberts *et al.*, 1994).

The structural details and function of ECM differ for different tissues and organs and although produced by the cells, ECM has been shown to influence cell behaviour. Turnover of ECM is therefore a crucial process for the control of cellular behaviour. It is essential in morphogenesis and tissue remodelling which occur during normal physiological conditions, such as embryonic development, wound healing and angiogenesis (Nagase and Woessner, 1999; Lund *et al.*, 1999). Misregulation of ECM components turnover, either by lack of degradation or excessive degradation, is linked to a series of pathological conditions (see Chapter II – Introduction to angiogenesis) (Halpert *et al.*, 1996; Konttinen *et al.*, 1999; Nelson *et al.*, 2000).

Modulation of cell-matrix interactions occurs through the action of proteolytic systems. Enzymes involved in these proteolytic systems regulate cell proliferation, differentiation and cell death. A number of proteinases are involved in the regulation of cell-matrix composition, the principal being the matrix-degrading proteinases known as Matrix Metalloproteinases (MMPs) or matrixins and the members of the family called a disintegrin and a metalloproteinase (ADAM; Nagase and Woessner, 1999; Brew *et al.*, 2000; Nagase and Brew, 2002). The role of these molecules is highly controlled under physiological conditions as disruption can lead to atherosclerosis, rheumatoid arthritis and cancer invasion and metastasis. Regulation of MMP activity is achieved by a plasma protein, α-2 macroglobulin in the fluid phase and by endogenous inhibitors of MMPs, the tissue inhibitors of metalloproteinases (TIMPs) in the tissue phase (Nagase and Brew, 2002). The latter group of inhibitors have attracted a lot of interest due to their potential use in therapeutics. TIMPs from several organisms have been studied to date and conserved characteristics among species have been identified, indicating potential evolutionary relationships. The study of one member of the TIMP family, *Drosophila melanogaster* TIMP (*d*TIMP) is the aim of this research. Analyses (sequence and structural) which provide information about its similarities and differences with its homologues from vertebrates and invertebrates are the means used to understand the evolutionary relationship among TIMPs.

4.2 Matrix metalloproteinases – MMPs

MMPs are multidomain proteins, actively involved in the remodeling of ECM; enzymes of this family can process almost every component of the extracellular matrix (Vu and Werb, 2000; Egeblad and Werb 2002; Puente *et al.*, 2003). A total of 24 genetically distinct MMPs have been identified in humans, while the total number of MMPs has risen to 33 (MEROPS database; Rawlings *et al.*, 2006; http://merops.sanger.ac.uk/). MMPs were initially categorised according to their substrate specificity, into collagenases, matrilysins, gelatinases, stromelysins, membrane-associated and others, while later a numerical classification was employed as the number of MMPs and the substrates they catalyse increased (Gross and Lapière, 1962; Park *et al.*, 2000; Uría and López-Otín, 2000). They are zinc-dependent endopeptidases and belong to a large family of proteinases, namely the Metzincin superfamily. Active members of this family are zinc–dependent and share the motif: HEXXHXXGXXH and a methionine turn (Met-turn; Bode *et al.*, 1993; Wei *et al.*, 2003). The domain organisation is conserved not only among different MMPs but among different species as well.

Conserved MMP domains include a signal peptide, the amino-terminal propeptide domain, the catalytic domain, the hinge region and the hemopexin-like domain at the carboxy-terminal (Figure 4.2; Massova *et al.*, 1998; Brew *et al.*, 2000). The signal peptide is required to direct secretion from the cell. The propeptide domain consists of around 90 amino acids, including a conserved cysteine. This cysteine residue has been shown to be involved in MMP regulation: it interacts with a zinc ion, present in the catalytic domain and keeps the MMPs in an inactive form, known as zymogen (Massova *et al.*, 1998). Removal of the propetide domain results in an active MMP form. The catalytic domain contains two zinc ions, one of which is essential for structural integrity, while the other is required for catalytic activity. Two to three calcium ions are also present in MMPs and have a role in the enzymatic activity and overall stability of the enzyme (Massova *et al.*, 1998). A linker or hinge region (of variable lengths) bridges the catalytic domain with the hemopexin domain at the carboxy-terminal of MMPs. The hemopexin-like domain is a highly conserved β-propeller structure. It has been shown to be involved in recognition, both for the binding of MMPs to their substrates as well as to

TIMPs, the main inhibitors of MMPs (Borden and Heller, 1997; Gomis-Ruth *et al.*, 1997).

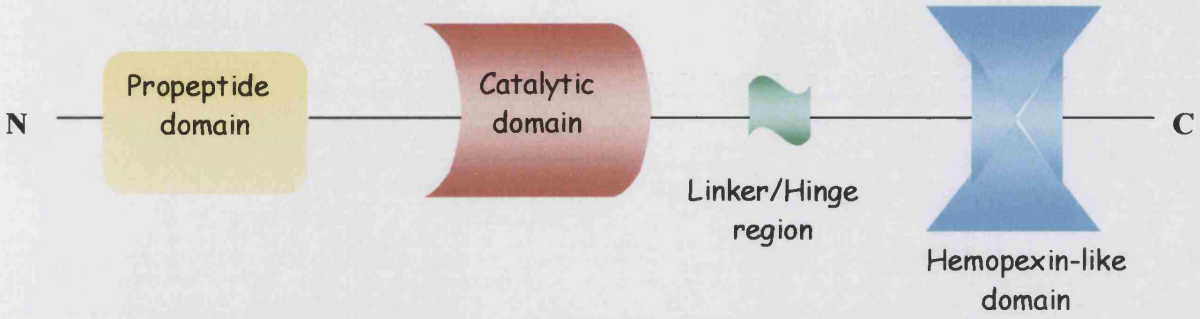


Figure 4.2: Domain organisation of MMPs. The conserved domains include the propeptide domain, the catalytic domain, the hinge region and the hemopexin-like domain. These domains are conserved not only among different types of MMPs, but among different species as well.

Additional domains, such as fibronectin-like repeats and hydrophobic repeats in the C-terminal domain have been identified in some of the members of the MMP family, increasing their structural diversity (Llano *et al.*, 2000).

Research on MMPs has resulted in the structural characterisation of individual domains as well as of the full-length protein. A full-length structure of porcine MMP-1 was determined at 2.5Å resolution by X-ray crystallography (PDB ID 1FBL; Li *et al.*, 1995). This structure illustrates the conserved domains, propeptide, catalytic domain, linker and hemopexin domain. Bound in the catalytic domain are two zinc ions (required for catalytic activity and structural integrity) and two calcium ions (required for the overall stability of the enzyme), while in the centre of the hemopexin four-bladed β-propeller structure a calcium ion is bound (Figure 4.3).

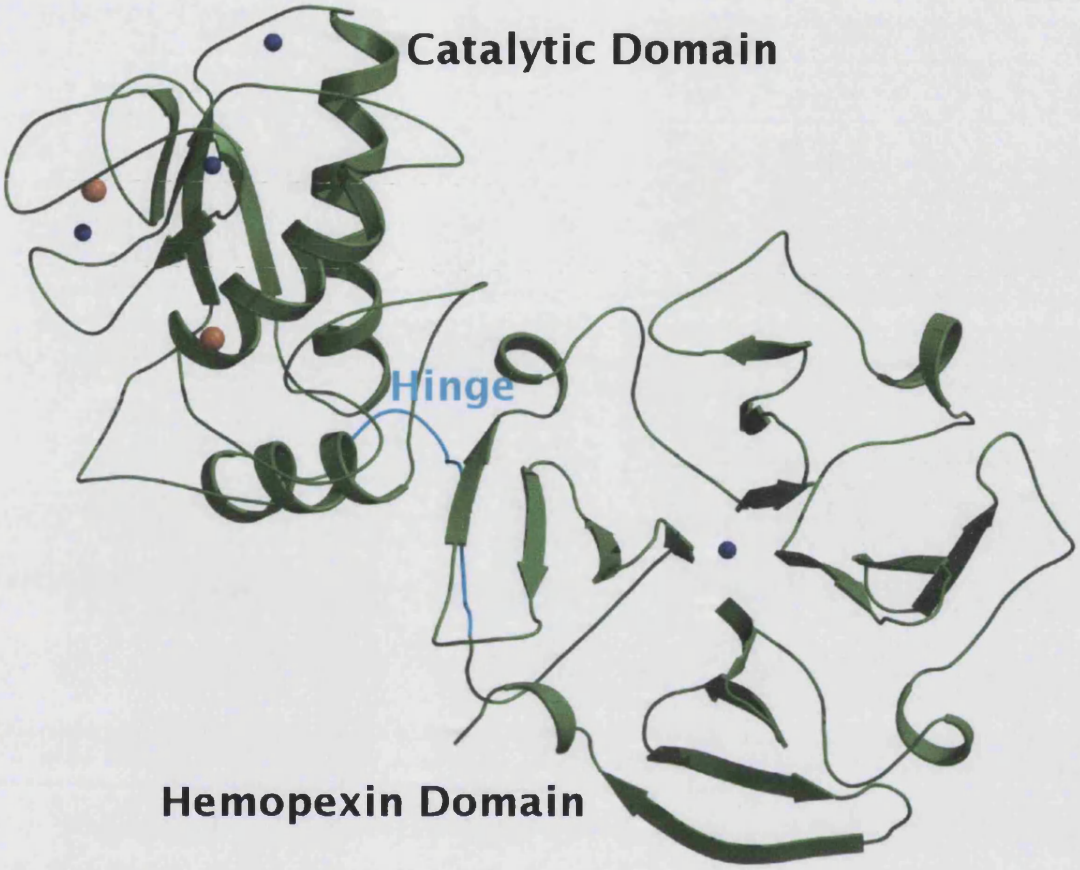


Figure 4.3: Crystal structure of porcine MMP-1 determined at 2.5Å resolution (PDB ID : 1FBL; Li *et al.*, 1995). Shown here are the catalytic domain, the hinge region (highlighted in cyan) and the hemopexin domain. Bound in the catalytic domain are two zinc ions (shown in orange) and three calcium ions (shown in blue). Bound in the centre of the hemopexin domain is a calcium ion (shown in blue).

As the main enzymes that catalyse the turnover of the extracellular matrix, MMPs are involved in physiological processes such as embryonic development, wound healing, angiogenesis and apoptosis (Werb, 1997; Murphy and Gavrilovic, 1999; Yu and Stamenkovic, 2000), as well as in pathological processes such as arthritis, atherosclerosis and tumour invasion and metastasis (Williamson *et al.*, 1990; Parks and Mecham, 1998).

Their implication in pathological processes has been extensively studied and has been directly associated with loss or imperfections in control of MMP activity (Sternlicht and Werb, 2001). Regulation of MMP activity is thus of utmost importance for the maintenance of a physiological environment and for avoidance of tissue destruction. Three major ways are employed for the regulation of MMP activity: restrained release of MMPs, secretion of MMPs in an inactive form (zymogen), or inhibition of MMP activity by α2-macroglobulin and by endogenous inhibitors, TIMPs. MMPs are normally secreted in an inactive form. Disruption of the bond between cysteines and zinc, which keeps MMPs in the inactive form, (known as the cysteine-switch) results in a partially active intermediate form of the enzyme, which is rendered completely active upon cleavage of the propetide domain, by autocatalysis (Van Wart and Birkedal-Hansen, 1990). Once activated MMPs can be regulated by inhibitors such as α2-macroglobulin in the fluid phase or by TIMPs in the tissue phase. TIMPs form tight 1:1 non-covalent complexes with MMPs inhibiting in this way their endopeptidase activity.

4.3 Tissue Inhibitors of Matrix Metalloproteinases – TIMPs

TIMPs are endogenous proteins that regulate MMP activity either directly by inhibiting their action or indirectly by modulating zymogen activation (Brew *et al.*, 2000; Wei *et al.*, 2003). TIMPs have been shown to be selectively expressed by spongiotrophoblastic cells, which separate the labyrinthine zone from the zone of giant cells. A border is formed in this way between foetal and maternal tissues (Henriet *et al.*, 1999). Some TIMPs have the ability to associate with the pericellular environment, as seen from the binding of TIMP-2 to cell surface MT1-MMP (membrane-type 1 MMP) which acts as a 'receptor' for proMMP-2 (progelatinase A) (Murphy and Knauper, 1997; Nagase, 1998). In mammals and higher organisms mature TIMPs are two-domain molecules organised in a single chain of around 180 amino acid residues (Williamson *et al.*, 1990).

TIMPs serve numerous biological functions, such as connective tissue re-modelling, cell growth promotion, matrix binding, induction of apoptosis as well as inhibition of active MMPs and pro-MMP activation. TIMPs stoichiometrically bind to MMPs to the ratio of

1:1 and irreversibly inactivate them. These enzymes are slow, tight-binding inhibitors and the TIMP-MMP complex is one of the tightest complexes known, with dissociation constant in the nanomolar range (Murphy and Willenbrock, 1995). Although some TIMPs have been shown to inhibit more than one MMP, there is generally specificity and selectivity in MMP inhibition by TIMPs.

TIMPs have also been shown to inhibit *in vitro* some members of the ADAM (a disintegrin and a metalloproteinase) family. Members of this family are multidomain proteins that consist of an N-terminal propeptide domain, a metalloproteinase domain, a disintegrin-like domain, a transmembrane domain and a cytoplasmic domain. They also contain the HEXXHXXGXXH motif for zinc binding in the catalytic domain (Black and White, 1998). One example of ADAM inhibition by TIMP is seen in *Drosophila melanogaster* (Hynes and Zhao, 2000), where seven ADAMs and ADAMTSs (TS – thrombospondin motif) genes have been identified. *d*TIMP is thought to be involved in the regulation of all or most of these metalloproteinases. Interaction of *d*TIMP with the ADAMs is thought to be mediated by the binding of *d*TIMP to the metalloproteinase domain of ADAMs. The inhibitory interactions have been shown to be of the subnanomolar range (Kashiwagi *et al.*, 2001).

TIMP structure is conserved among species: all TIMPs have a 23-29 amino acid residue leader sequence that is cleaved to produce the mature TIMP protein (Pohar *et al.*, 1999). A characteristic feature of TIMPs is the twelve cysteines, all of which are involved in disulphide bonds (Figure 4.4). Two of these cysteines are present in the very beginning of the N-terminal domain and form the C-x-C motif, which is characteristic of all mature TIMPs. The first cysteine is of major importance, as it coordinates the catalytic zinc of the MMP that each TIMP inhibits. Mammalian TIMPs have an N-terminal domain of about 125 amino acid residues and a smaller C-terminal domain of about 65 amino acid residues. Each domain is stabilised by three disulphide bonds (Figure 4.5). Of the two domains, the N-terminal domain of TIMPs has been shown to be solely responsible for their inhibitory activity (Williamson *et al.*, 1990). The role of the C-terminal domain is

not yet clear, but it is thought to assist in the orientation of the inhibitory end when binding to the MMPs (Langton *et al.*, 1998).

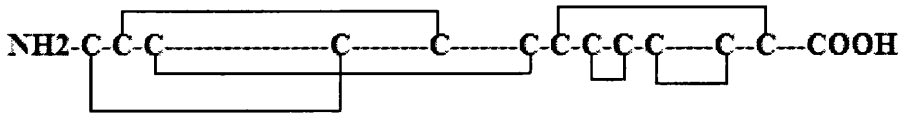


Figure 4.4: Characteristic motif of mature TIMPs, where the leader peptide has been cleaved. Shown here are the twelve cysteines, conserved in vertebrates and all involved in disulphide bonds (Figure adapted from: Prosite; Hulo *et al.,* 2006).

The N-terminal domain of both vertebrate and invertebrate TIMPs contains an oligosaccharide/oligonucleotide binding fold, and thus falls in the OB-fold category (Williamson *et al.*, 1994; Gomis-Ruth *et al.*, 1997; Tuuttila *et al.*, 1998). It consists of a five-stranded β-barrel, with a Greek key topology, and two α-helices; it is relatively rigid and shows almost no movement on binding to MMPs. Characteristic of TIMPs is also the presence of an AB loop in the N-terminal domain (Figure 4.5; PDB ID 1UEA; Gomis-Ruth *et al.*, 1997). No functional significance has been assigned to the AB loop although it seems to differ in length and arrangement in different TIMPs. One such example is the more extended AB loop of human TIMP2 in contrast to human TIMP1. Although human TIMP2 has been shown to form stronger complexes than human TIMP1, it has not been attributed to the extended AB loop (Brew *et al.*, 2000).

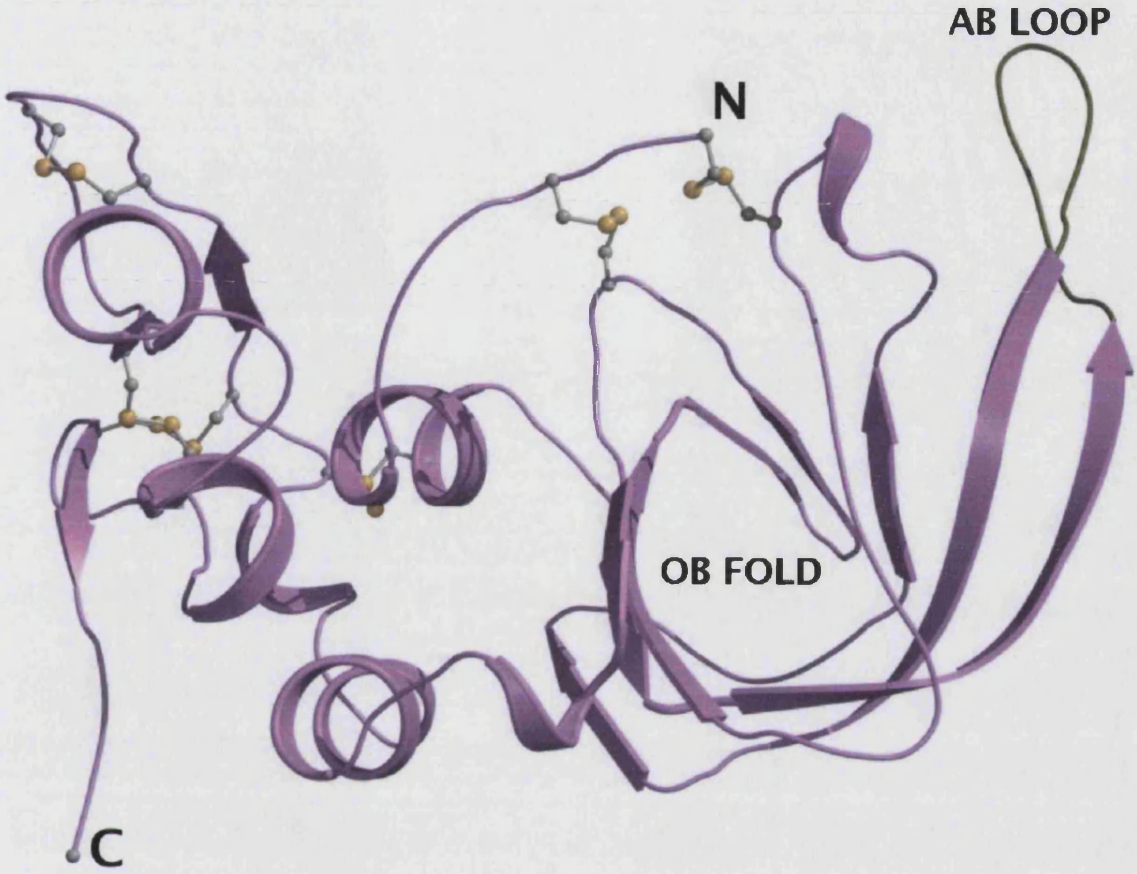


Figure 4.5: Structure of human TIMP1 (PDB ID: 1UEA; Gomis-Ruth *et al.*, 1997). Shown in ball-and-stick are the disulphide bonds, conserved in TIMPs; three in the N-terminal domain and three in the C-terminal domain. Also labelled are the OB fold and the AB loop, characteristic features of TIMPs.

To date, four genetically distinct TIMPs have been cloned and characterised in humans (TIMP1 to TIMP4), with crystal structures available for TIMP1 and TIMP2. The number of TIMPs differs amongst species with most mammals having either two or four TIMPs, while most invertebrates seem to have a single TIMP copy (e.g. *Drosophila melanogaster*). Some nematodes also have a single TIMP (e.g. *Necator Americanus*), while others have more TIMP alleles (e.g. *Hydra vulgaris* TIMP1-2 and *Ascaris lumbricoides* TIMP1-5).

*4.3.1 TIMP from Drosophila melanogaster - d*TIMP

TIMPs from different species have been studied, mainly those from human and bovine. A TIMP that has recently attracted interest is the TIMP from *Drosophila melanogaster* (*d*TIMP), which seems to have some unique features, while retaining the basic characteristics of TIMPs. Only one TIMP copy is present in *Drosophila melanogaster* and is responsible for the inhibition of the two MMPs (*Dm*MMP1 and *Dm*MMP2) as well as for the control of the activity of seven ADAMs also present in *Drosophila*. Mutational studies where *d*TIMP was knocked-out, resulted in inflated wings, bloated gut and death (Godenschwege *et al.*, 2000). This is most likely the result of unregulated MMP and ADAM activity, as *d*TIMP is the sole enzyme responsible for their regulation.

*d*TIMP is a single protein sequence, consisting of 210 amino acids, 12 of which are cysteines involved in disulphide bonds. The full-length molecule greatly resembles its mammalian homologue: the C-terminal region is tightly packed against the N-terminal OB-fold. The characteristic motif of C-x-C in the N-terminal domain of mature TIMPs is also present in *d*TIMP, after a leader sequence of 27 amino acids. Interestingly, *d*TIMP appears to lack two cysteines from the N-terminal domain, having thus two, instead of three disulphide bonds in the N-terminal domain. In contrast to its mammalian homologues, *d*TIMP has two extra cysteines in the C-terminal domain, having thus an extra disulphide bond (four disulphide bonds in the C-terminal domain of *d*TIMP; Figure 4.8; Brew *et al.*, 2000). The inhibitory activity of *d*TIMP appears to originate from the N-terminal domain, a characteristic that is conserved among TIMPs (Wei *et al.*, 2003). The functional importance of the C-terminal domain (consists of around 65 amino acids) is not yet clear, as its removal did not appear to affect the inhibitory activity of *d*TIMP, as seen from experiments where *d*NTIMP was tested for its inhibitory activity against human MMPs, *Dm*MMPs and human TACE (tumour necrosis factor-α-converting enzyme). Experimental results revealed that *d*TIMP is a potent inhibitor of MMPs (from human and *Drosophila melanogaster*) as well as of TACE (human and *Drosophila melanogaster*) (Wei *et al.*, 2003). Conservation of the N-terminal inhibitory activity in the absence of one disulphide bond might indicate that the third disulphide bond observed in mammalian TIMPs is an evolutionary acquisition rather than a conserved TIMP

feature. Wei and co-workers (Wei *et al.*, 2003), suggested that *d*TIMP is most similar to mammalian TIMP3, suggesting that the latter is more closely related to the ancestral TIMP gene than the rest of mammalian TIMP classes. Hence, according to these findings, TIMPs classes 1, 2 and 4 are more recent and have probably diverged through gene duplication or mutational changes (Wei *et al.*, 2003). Brew and co-workers (Brew *et al.*, 2000), proposed that TIMP1 has undergone the lowest rate of evolutionary change, making it the most proximal to the ancestral TIMP, from which the rest of the human TIMPs originated (Brew *et al.*, 2000). It is evident that the findings of the two groups are not in agreement, indicating that further research in *d*TIMP and its proximity to one of the human TIMPs would be required, in order to understand the ancestry and evolutionary relationship among TIMPs.

In order to obtain more information about ancestry and understand the evolutionary relationship of TIMPs, multiple sequence alignments and phylogenetic tree analyses were carried out. The structures, where available, and models (created for TIMPs whose structure has not been reported yet) were aligned and the root mean square deviation was calculated, as an indication of similarity or divergence. For that purpose, TIMP sequences from vertebrates as well as invertebrates were used in this study; all sequences used, are provided and discussed later in 'Results and Discussion' section (Table 4.2).

4.4 Bioinformatic analysis

Protein structure prediction is a widely used method that employs bioinformatic tools and aims at the determination of the three-dimensional structure of proteins, from their amino acid sequence only (primary structure). There are two types of protein structure prediction: *Ab initio* protein modeling and homology modeling. The first method aims at determining the three dimensional structure of a protein based on physical properties. This method involves algorithms like Rosetta and Monte Carlo, while Ramachandran plot can also be used for *ab initio* modeling. *Ab initio* protein modeling can be considered to consist of two steps: the generation of a scoring (energy) function that distinguishes between native or native-like structures, and a search method to explore the

conformational space, which is an energy function that should account for all interactions between all pairs of atoms in the polypeptide chain.

Homology modeling for protein structure prediction aims at determining the structure of a protein based on previously determined three-dimensional structures of homologous proteins. This method is based on the evolutionary conservation of the fold of a protein. A sequence alignment of the protein of interest with homologous proteins, whose three-dimensional structures are available, is the first step employed by this method, so as to find the closest match that will be used as a template. Hence, it is a more accurate method for protein structure prediction than *ab initio* modeling, but requires that structures of homologous proteins are available. Homology modeling was the method used for generating a model for *d*TIMP, as structures of homologous proteins are available. Bioinformatic structure prediction can be considered to consist of eight stages: literature search, sequence search, motif search, sequence alignment, secondary structure prediction, tertiary structure prediction, model refinement and finally validation of the generated model (Figure 4.6).

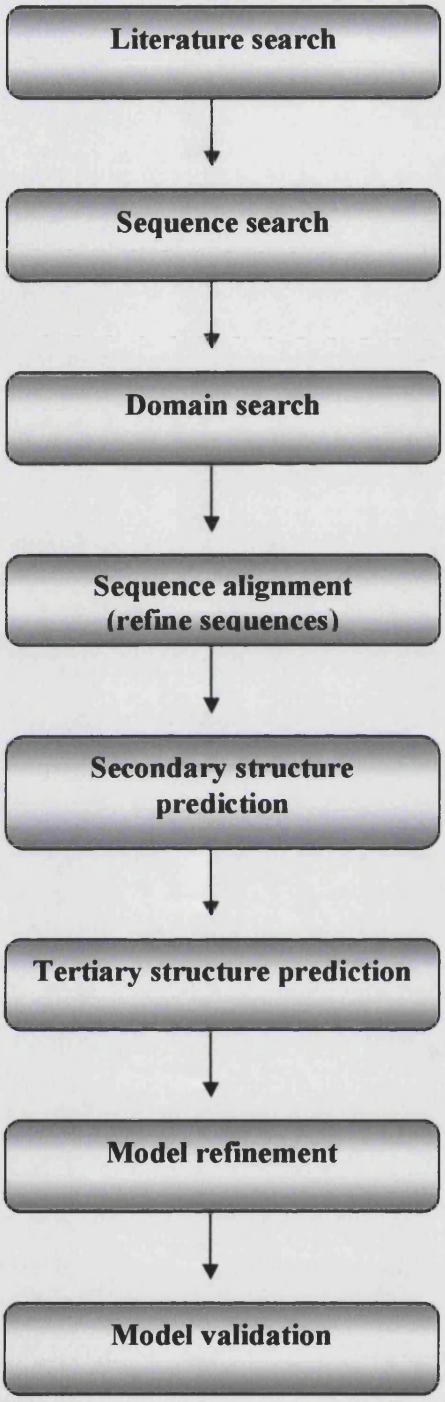


Figure 4.6: Flowchart representation of the steps involved in bioinformatics analysis for protein structure prediction.

4.4.1 Literature search

Literature search is fundamental for structure prediction, as it provides information about the protein of interest, the structure and function of homologous proteins and any previous efforts for predicting the structure of the protein of interest. The pubmed database (http://www.ncbi.nlm.nih.gov/) allows for an extensive search on literature with links on the national center for biotechnology information, the national library of medicine and the national institutes of health. It is essential however to be critical on the information retrieved as literature may contain errors or the protein of interest might not be a typical member of well-characterised families. Hence an extensive literature search should be the first step in understanding the characteristics of the protein of interest and identifying homologues that can be used in further steps of the protein structure prediction.

4.4.2 Sequence search

Sequence search is an essential step in identifying protein sequences 'related' to the sequence of the protein of interest, that have been previously characterised and can be used as templates for building the model of the query. A tool that is most commonly used for sequence search is BLAST (Altschul *et al.*, 1990), a web-based program accessible via NCBI. BLAST (Altschul *et al.*, 1990) searches for local, ungapped alignments, which might prove to be a disadvantage, as low sequence similarity does not necessarily mean low structural similarity. Another web-based program which is frequently used for sequence search is FASTA (Pearson, 1990; Pearson and Lipman, 1998). This is a more sensitive web-based program that looks for small regions of similarity between the query and other known protein sequences. This type of search yields a list of sequences that are 'good matches' for the query and are presented in descending order. Apart from the retrieved sequences, each program provides a list of parameters including the length of the retrieved sequences, the similarity with the query sequence and the expectation value. The expectation value (also known as the E value) represents the average amount one expects something to re-occur in the database where it was originally searched. Hence, the lower the E value, the more significant the score and the better the indication of the proximity of the sequences retrieved by the database. More information on the protein of

interest can be obtained *via* the analysis of its amino acid sequence in the Protparam service (Gasteiger *et al.*, 2005), available in the Expert Protein Analysis System server (www.expasy.org). Such an analysis can provide information on the molecular weight, the isoelectric point (pI), the extinction coefficient and the instability index of the protein of interest, which can prove to be very helpful in subsequent experimental steps.

4.4.3 Domain search

Domain search refers to the search for short, tightly defined patterns of polypeptide sequences that might be indicative of protein function. It produces a list of known domains, families and sites which are associated with protein function that have been identified in the query sequence. The most commonly used tool for motif search is Prosite (Hulo *et al.*, 2006). Prosite allows the identification of possible domains and characteristic features (such as conserved disulphide bonds) according to the amino acid sequence of the protein of interest and provides literature documentation available on the identified characteristics. Hence, findings obtained from Prosite provide information for the protein of interest, by examination of members of the same family, both in structural and functional level. A web-based server which allows the prediction of domains based on the deposited sequence of the protein of interest is Meta – DP (Meta Domain Prediction), which has the advantage to be coupled to different prediction servers. This allows the prediction of domains in the query sequence, even if no information for a homologue to the query protein is available (Saini and Fischer, 2005).

4.4.4 Sequence alignment

Preliminary sequence alignments are provided by BLAST, which aligns each of the results retrieved against the query. Multiple sequence alignments allow the simultaneous alignment of more than one sequence. Commonly used alignment programs are ClustalW (Thompson *et al,*. 1994) and Tcoffee (Notredame *et al.*, 2000; Poirot *et al.*, 2003). This is advantageous in that it highlights conserved residues, which in turn might prove to be significant for the structure and function of the protein of interest. Conserved regions identified in this way can be used for more sensitive analyses, using either a domain

search or a PSI-BLAST (Altschul *et al.*, 1997) search, which provides additional information on profiles identified in the deposited sequences.

4.4.5 Secondary structure prediction

Secondary structure prediction is the identification of secondary structure elements, which in turn can provide information on the topology of the protein of interest. Several web-based servers are available for the prediction of the secondary structure of a protein, where the amino acid sequence of the query is submitted and the results are usually mailed back to the user. Such programs use either sequence alignments or available PDB files in order to identify similarities between the query sequence and a protein (usually homologous) which has been already characterised. PHD is one such web-based program with an accuracy of around 71.4% for proteins which have at least one known homologue (Rost *et al.*, 1994). Another web-based server for protein secondary structure prediction is JPRED, which was developed by the University of Dundee and has an accuracy of around 76.4% (Cuff *et al.*, 1998). PROF is software for protein secondary structure prediction which can be accessed online, or downloaded for private use. It was developed from the University of Aberystwyth and requires PSI-BLAST and ClustalW in order to operate (Thompson *et al.*, 1994). A commonly used web-based server for protein secondary structure prediction is PSIPRED, using a PSI-BLAST output as its input file (Jones, 1999). This server is advantageous in that it allows the prediction of protein secondary structure, the transmembrane topology (if applicable) and the fold recognition, based on the amino acid sequence. SAM-T99 uses Hidden Markov Models (HMM), a statistical method which uses available PDB structures in order to predict the secondary structure of the deposited query sequence (Karplus *et al.*, 1999). SSpro is another web-based server which uses homology analysis for the prediction of protein secondary structure (Cheng *et al.*, 2005). An extensive list of available programs for protein secondary structure prediction is available on the Meta-Predict Protein server (Eyrich and Rost, 2003), which also provides a model of the query sequence, if requested, and on a web-page developed by the University of York for analysis and modeling of proteins (http://www.ysbl.york.ac.uk/~tom/structure.html#2D-prediction). Considering all the information that can be retrieved by such analyses, it is possible to say that secondary

structure prediction gives a "first image" of the tertiary structure and the overall folding of the query protein. Further information about the tertiary structure and a prediction of the structure of the protein of interest, can be obtained by tertiary structure prediction.

4.4.6 Tertiary structure prediction

Tertiary structure prediction refers to the modeling of the query protein, using all the data collected in previous steps. The most commonly used program for molecular modeling is the SWISS-MODEL server (Schwede *et al.*, 2003), an automated comparative modeling server. SWISS-MODEL (Schwede *et al.*, 2003) hides all the technical details and yields a three-dimensional model of the query. The choices in SWISS-MODEL (Schwede *et al.*, 2003) include the first approach mode, where the query sequence is deposited and the structure is predicted using known structures as templates. The input can also be a sequence alignment (alignment interface) where the user can choose the best match to the query and use it as a template. Finally, project (optimise) mode allows the deposition of the sequence and the superimposed structures to be used as templates. A three-dimensional model of the query is then built according to the structures chosen as templates. Meta-protein predict (MPP; Eyrich and Rost, 2003) is a web-based server for the creation of a three-dimensional model, based on the amino acid sequence of the protein of interest. MPP allows access to other web-based servers which provide information on the secondary structure prediction of the protein of interest.

Once a model for the protein of interest has been generated, it is usually a good idea to check the structural features of proteins that belong to the same 'class'. Two web-based servers which are commonly used for retrieving information on proteins according to their 'class' are the structural classification of proteins – SCOP (Murzin *et al.*, 1995) and the classification according to class, architecture, topology and homologous superfamily – CATH (Orengo *et al.*, 1997) databases, which provide information on the classification of protein patterns. Both can be accessed *via* TOPS (Michalopoulos *et al.*, 2004), one of the most commonly used topology databases. TOPS (Michalopoulos *et al.*, 2004) produces a first estimation of the pattern and topology of the protein of interest as well as a cartoon representation of the organisation of the domains in query protein. The division

of the primary structure into secondary structure elements can be then compared to the protein secondary structure prediction outputs and the model generated for the protein of interest.

4.4.7 Model refinement

Once the three-dimensional model of the protein of interest is built, refinement is essential, in order to ensure the model is meaningful, informative and devoid of any steric hindrances. Model refinement refers to the use of physico-chemical modeling methods for the correction of mistakes in the model. Most common errors in modeled structures involve distorted geometry (bond angles and bond lengths) and overlapping atoms, which can be corrected during model refinement. Input files from CNS (Brünger *et al.*, 1998) can be used for refining the model, by removing the crystal card from the input files, which then allows the use of the coordinate file and set parameters for refining. Most commonly used steps for model refinement include energy minimisation and molecular dynamics. After the model is refined, it should also be checked manually, in one of the graphics programs, such as 'O' (Jones *et al.*, 1991) and 'Coot' (Emsley and Cowtan, 2004).

4.4.8 Model validation

As with structure validation, model validation is the fine-tuning step, where small errors are corrected, using the same methods that are used for three dimensional structures before their deposition to the protein data bank (PDB; Berman *et al.*, 2000). PROCHECK (Laskowski *et al.*, 1993) is the most commonly used program for validating protein structures and models. Once a model has been refined and validated, it can be deposited to SWISS MODEL repository, a database of models, currently counting 1,341,793 deposited models.

4.5 Results and discussion

Literature search in pubmed provided background information on *d*TIMP, its proposed characteristics and features. Once sufficient information was obtained, the next step was to retrieve the amino acid sequence and analyse it. The amino acid sequence of *d*TIMP was retrieved from the expasy website, with primary accession number Q9VH14. *d*TIMP consists of 210 amino acids, while the mature protein, starting with the C-x-C motif, consists of 183 amino acids. Mature *d*TIMP has a molecular weight of ~21kDa and a theoretical pI of 9.6. According to its instability index (II), calculated by Protparam, based on the amino acid sequence, *d*TIMP can be classified as unstable, with an II value of 55.5 (Gasteiger *et al.*, 2005). A search against the PDB, using BLAST as a tool, yielded a series of matches, with the best 'hits' ranked first in the results list (best 'hits' for *d*TIMP are provided in Table 4.1).

Table 4.1: BLAST search against the Protein Data Bank.

Results from the BLAST search	Expectation Value
2E2D – Flexibility and variability of TIMP binding: X-ray structure of the complex between collagenase-3/MMP-13 and TIMP-2 (Maskos *et al.*, 2007) *Homo sapiens and Bos Taurus*	$5e^{-10}$
1BUV – Crystal structure of MT1-MMP-TIMP-2 complex (Fernandez-Catalan *et al.*, 1998) *Homo sapiens* and *Bos taurus*	$5e^{-10}$
1UEA – MMP-3/TIMP-1 complex (Bode *et al.*, 1997) *Homo sapiens*	$2e^{-8}$
2TMP – N-terminal domain of tissue inhibitor of metalloproteinase N-TIMP2 (Muskett *et al.*, 1998) *Homo sapiens*	$6e^{-7}$
2J0T – Crystal structure of the catalytic domain of MMP1 in complex with the inhibitory domain of TIMP1. (Iyer *et al.*, 2007) *Homo sapiens*	$3e^{-6}$

Note: The sequences with the highest expectation value are given in this table. Not all results of complexes of MMPs with TIMP2 are provided, as the expectation values were all in the same range (e^{-10}).

Results obtained include structures that have been determined by NMR and/or X-ray crystallography. The best 'hit' obtained was the structure of the complex of MMP13-TIMP2 (PDB ID 2E2D; Maskos *et al.*, 2007), with an expectation value of $5e^{-10}$. Several hits with an expectation value of e^{-10} were retrieved for complexes including *h*TIMP2.

A Prosite analysis of *d*TIMP (Hulo *et al.*, 2006) revealed a netrin module (NTR module), which is typically found in the C-terminal domains of netrins, as well as in the N-terminal domains of TIMPs. A typical feature of the NTR module are six cysteines, (conserved in proteins that contain this particular module), which most probably form disulphide bonds, just as observed in TIMPs. Domain analysis using the Meta-DP server (Saini and Fischer, 2005) also indicated an NTR domain in the amino acid sequence of *d*TIMP, just as previously obtained by Prosite (Hulo *et al.*, 2006).

Based on the best 'hits' obtained from BLAST, TIMPs from vertebrate and invertebrate sources were selected for the following steps of the bioinformatics analysis and modeling process. Selected TIMPs are provided below, in Table 4.2.

Table 4.2: TIMPs included in the analysis and building of the *d*TIMP model.

Organism	Common Name	TIMP class	Full length	Accession Number
Drosophila melanogaster	Drosophila	TIMP	210	**Q9VH14**
TIMPs from vertebrates				
Homo sapiens	Human	TIMP1	207	**P01033**
		TIMP2	220	**P16035**
		TIMP3	211	**P35625**
		TIMP4	224	**Q99727**
Mus musculus	Mouse	TIMP1	205	**P12032**
		TIMP2	220	**P25785**
		TIMP3	211	**P39876**
		TIMP4	224	**Q9JHB3**
Rattus norvegicus	Rat	TIMP1	217	**P30120**
		TIMP2	220	**P30121**
		TIMP3	211	**P48032**
		TIMP4	224	**P81556**
Scyliorhinus torazame	Cloudy catshark	TIMP3	214	**Q9W6B4**
Xenopus laevis	African clawed frog	TIMP3	214	**O73746**
TIMPs from invertebrates				
Pristionchus pacificus		TIMP	136	-
Necator americanus	Strongylida	TIMP	140	-
Hydra TIMP1		TIMP1	201	-
Hydra TIMP2		TIMP2	201	-

Sources of TIMPs used in multiple sequence alignments. The mammalian sequences have been retrieved from the expasy website while nematode sequences were provided by our collaborators in Florida (Dr. Keith Brew, Florida Atlantic University).

TIMPs from *homo sapiens*, *mus musculus* and *rattus norvegicus* were selected as representatives of mammalian TIMPs which consist of four TIMP classes. *Scyliorhius torazame* and *Xenopus laevis* were also selected for the group of vertebrate TIMPs to be compared to *d*TIMP. As *d*TIMP is expected to represent a more ancestral TIMP form, invertebrate TIMPs were also included in the analysis. Invertebrate TIMPs examined, include *Pristionchus pacificus*, which is a free-living bacterivore nematode and *Necator*

americanus which is a type of hookworm that severely affects humans, leading to severe blood loss, cardiac and pulmonary complications. Also used in this analysis was *Hydra*, one of the most well known members of the hydrozoan species. Comparative analysis of TIMP representatives from vertebrate and invertebrate species against *d*TIMP would provide information about their proximity and probable evolutionary relationship. Using all the results from secondary structure prediction, the available structures and homology models, tertiary structure prediction tools were employed to generate a model for *d*TIMP (Figure 4.7). The model generated was energy minimised (using CNS; Brünger *et al.*, 1998) and validated in order to ensure it was devoid of any steric hindrances.

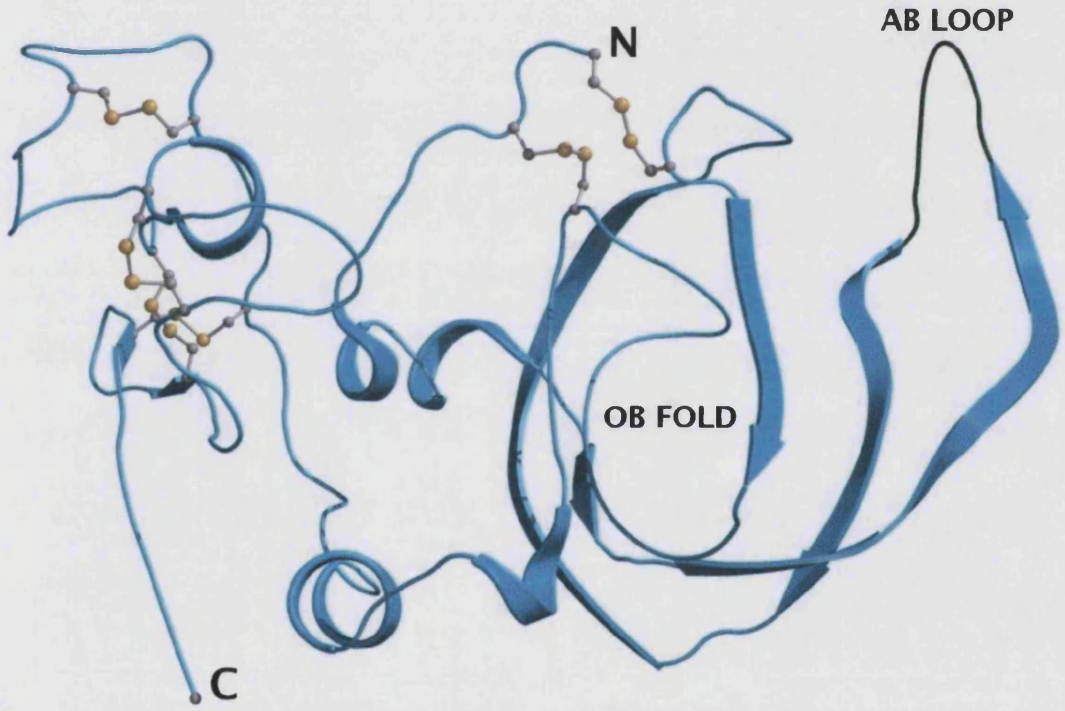


Figure 4.7: Model generated for *d*TIMP. Shown in ball-and-stick are the disulphide bonds, conserved in all TIMPs. The organisation of disulphide bonds in *d*TIMP differs from its mammalian homologues. Instead of three disulphide bonds in each domain, *d*TIMP has one less in the N-terminal domain, 'compensated' by an extra bond in the C-terminal domain. Also labeled are the OB fold and the AB loop, characteristic features of all TIMPs.

The homology model generated for *d*TIMP resembles its mammalian homologues. The N-terminal domain contains an OB fold and an AB loop, just as seen from the three dimensional structures for *h*TIMP1 and *h*TIMP2. The C-terminal domain appears to be less ordered, with several loops connecting the secondary structural elements of this domain. Whilst six disulphide bonds are observed in *d*TIMP their organisation and dispersal between the domains is different.

To identify the proximity or distance in structural/model level the root mean square deviation was calculated for all aligned structures and models. Due to differences in TIMPs from different sources, two groups have been created and analysed against *d*TIMP: *d*TIMP against vertebrate TIMPs and *d*TIMP against invertebrate TIMPs. Two insect TIMPs have also been studied and analysed against *d*TIMP.

4.5.1 dTIMP against vertebrate TIMPs

The amino acid sequence of *d*TIMP was aligned against its vertebrate homologues, using a multiple sequence alignment tool, in T-coffee, a web-based server (Notredame *et al.*, 2000; Poirot *et al.*, 2003; Figure 4.8). The alignment obtained was then used as an input for Boxshade, a web-based server that identifies conserved regions on multiple sequences alignments and colours them accordingly (Figure 4.8).

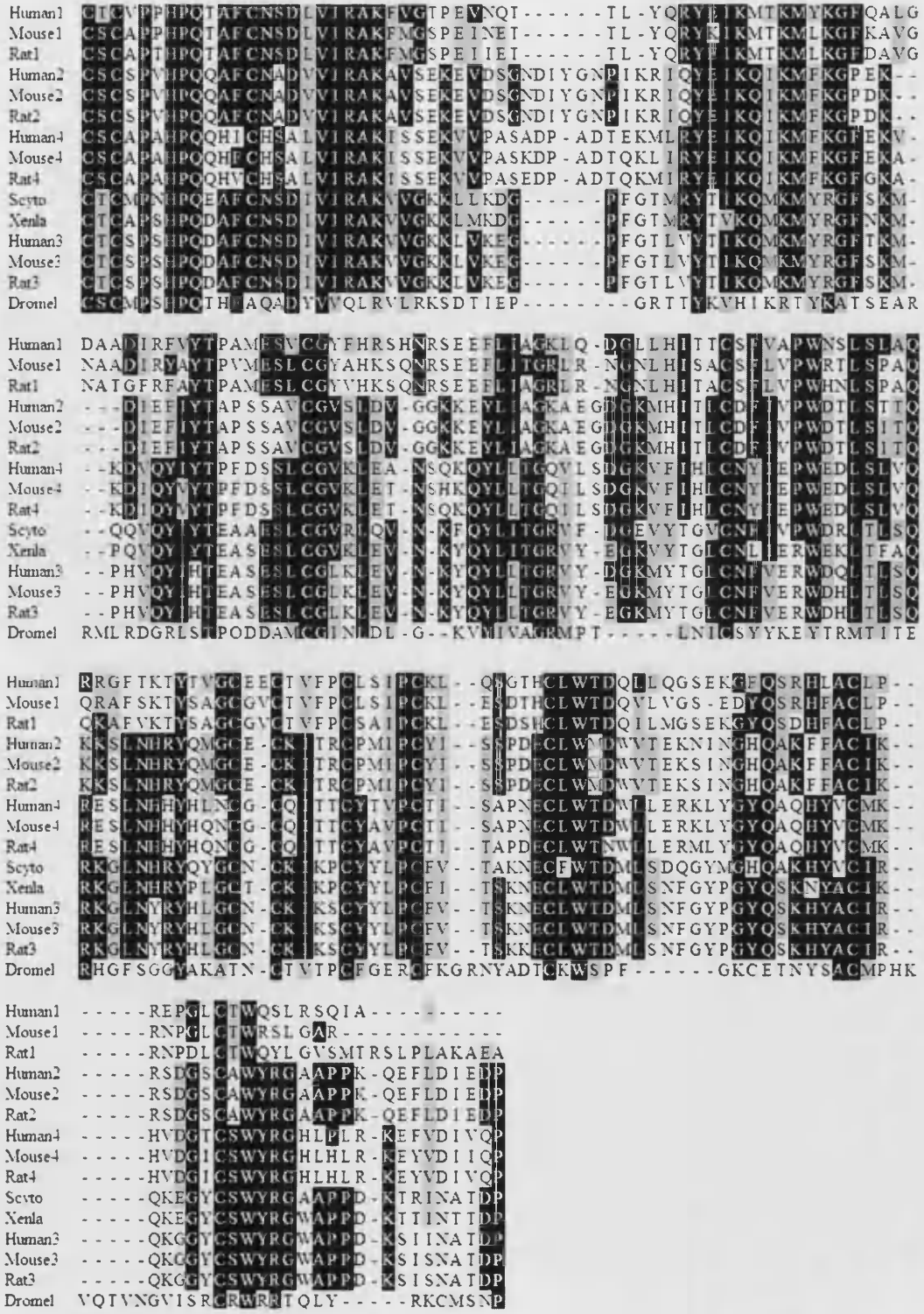


Figure 4.8: Full length sequence alignment of *d*TIMP against vertebrate TIMPs. Conserved and equivalent residues are highlighted here using Boxshade. *d*TIMP only shares a few of the conserved residues of vertebrate TIMPs, most importantly the conserved cysteines. Multiple sequence alignment was generated using the T-coffee server (Notredame *et al.*, 2000; Poirot *et al.*, 2003).

Several areas of black background, as seen in the sequence alignment, are shared among vertebrate TIMPs, indicating conserved residues. The overall sequence homology of the sequences aligned was 51.3%. Although the percentile sequence homology is high, *d*TIMP does not share most of the conserved regions of vertebrate TIMPs, as seen from the Boxshade colouring of the multiple sequence alignment.

Phylogenetic tree analysis was performed to identify which class of vertebrate TIMPs is most proximal to the ancestral TIMP form, represented by *d*TIMP. The tool employed for generating the phylogenetic tree, in the form of a phylogram, was the Genebee server (Brodsky *et al.,* 1992), a web-based tool for multiple sequence alignments and phylogenetic tree generation (Figure 4.9).

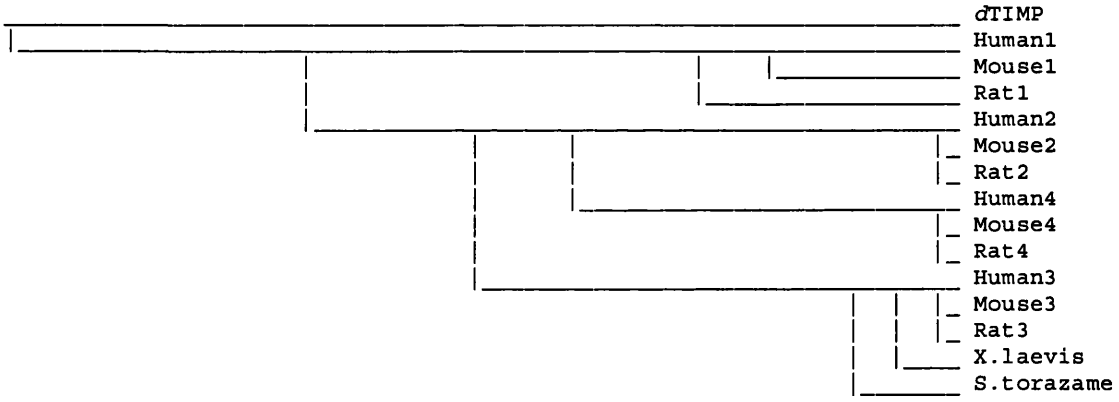


Figure 4.9: Phylogenetic tree analysis of *d*TIMP and vertebrate TIMPs. As the most ancestral TIMP form, *d*TIMP forms the root of this tree while vertebrate TIMPs occupy subsequent branches. Most proximal to *d*TIMP appears to be *h*TIMP1 while the class of TIMP3 appears to be more distant to *d*TIMP.

As seen in Figure 4.9, *d*TIMP does not "cluster" with any of the vertebrate TIMPs and as representative of the most ancestral form *d*TIMP is the root of the phylogenetic tree. It has been previously suggested that *h*TIMP1 has undergone the lowest rate of evolutionary change, and is thus the most proximal of all TIMPs to the ancestral form

(Brew *et al.*, 2000). This is in agreement with the results obtained from the phylogenetic tree analysis: *h*TIMP1 appears to be closer to *d*TIMP than any other vertebrate TIMPs, as they both appear to form the root of the tree. This in not however the case for the other vertebrate TIMP 1 members; mouse TIMP1 and rat TIMP1 appear to occur later, 'branching off' from *h*TIMP1. *h*TIMP2 is the next class of TIMPs most proximal to *d*TIMP as it occurs directly after *h*TIMP1. Following are *h*TIMP3 and *h*TIMP4, which occupy 'later' branches in the phylogenetic tree, indicating thus that they probably evolved later. Of all vertebrate TIMPs used in this analysis, *h*TIMPs appear to be most proximal to *d*TIMP, rather than TIMPs from mouse (*m*TIMPs) and rat (*r*TIMPs).

In order to identify the proximity or distance of the vertebrate TIMPs to *d*TIMP, structures (where available) and homology models, were aligned and rmsd calculation was used as measure of proximity or distance (Table 4.3; the better the structures align the lower the rmsd). As mentioned above there are limited structures of TIMPs available, so for vertebrate TIMPs with no structure reported, homology models were produced and aligned against *d*TIMP. All models generated were energy minimised using CNS (Brünger *et al.*, 1998) in order to be consistent and compare meaningful models.

Table 4.3: Structural comparison of *d*TIMP against vertebrate TIMPs.

Structures/models aligned	RMSD (Å)
*d*TIMP–*h*TIMP1	1.6
*d*TIMP–*m*TIMP1	1.6
*d*TIMP–*r*TIMP1	1.6
*d*TIMP–*h*TIMP2	1.4
*d*TIMP–*m*TIMP2	1.5
*d*TIMP–*r*TIMP2	1.5
*d*TIMP–*h*TIMP3	1.5
*d*TIMP–*m*TIMP3	1.4
*d*TIMP–*r*TIMP3	1.4
*d*TIMP–*X.laevis*	1.4
*d*TIMP– *S.torazame*	1.5
*d*TIMP–*h*TIMP4	1.5
*d*TIMP–*m*TIMP4	1.4
*d*TIMP–*r*TIMP4	1.5

RMSD is calculated for backbone (C^{α}) only using the program SPDBV (Guex and Peitsch, 1997).

Vertebrate TIMPs shared an overall sequence homology of 51.3% with *d*TIMP. Structural alignment of different TIMP classes against *d*TIMP shows little variation, as seen from the rmsd. Alignment of vertebrate TIMP1 class against *d*TIMP yielded an rmsd in the range of 1.6Å, indicating the differences among the structures and homology models that were aligned, although the homology percentage of this vertebrate class against *d*TIMP was 54.4%. It is thus evident that sequence homology and phylogenetic

proximity cannot be used as a 'representation' or 'indication' of possible structural similarity.

Vertebrate TIMP2 class has much higher sequence homology to *d*TIMP, in the range of 62%, indicating its proximity to *d*TIMP, in the sequence level. Structures and models of this vertebrate class yielded an rmsd in the range of 1.4-1.5Å when aligned against *d*TIMP. This rmsd range is not significantly lower than that obtained for vertebrate TIMP1 against *d*TIMP, even though the sequence homology is significantly high.

Similar are the findings from the alignment of members of the vertebrate TIMP3 class (including *X.laevis* and *S.torazame*) against *d*TIMP: sequence homology is even higher in this case, reaching the range of 72%. Structural alignment of vertebrate TIMP3 against *d*TIMP yields an rmsd in the range of 1.4Å, slightly lower than those obtained previously, but not significantly different.

Sequence homology of vertebrate TIMP4 against dTIMP is in the range of 64%, around the same range as that of vertebrate TIMP2. Structural similarity, as expressed by rmsd, for vertebrate TIMP2 against *d*TIMP is in the range of 1.5Å, again in the same range as that observed for members of the vertebrate TIMP2 class.

Based on these findings, it is possible to say that, although closer in the phylogenetic tree, TIMP1 is not most proximal to *d*TIMP in sequence level. This appears to be vertebrate TIMP3, with a very high sequence homology percent, which is not however supported by the phylogenetic tree findings. This is rather interesting as sequence homology would be expected to agree with phylogenetic proximity. It is not thus yet clear whether vertebrate TIMP1 or TIMP3 is the most proximal vertebrate class to *d*TIMP. Vertebrate TIMP 2 and TIMP4 on the other hand appear to be around the same ranges, both in sequence homology and structural similarity against *d*TIMP. It should be mentioned that as models were used for the structural alignment, further crystallographic studies need to be carried out in order to authenticate the results of the present study.

4.5.2 dTIMP against invertebrate TIMPs

TIMPs have also been identified in invertebrates with no structures reported to date. Some invertebrates have a single TIMP copy while others have more TIMP alleles (e.g *Ascaris*). Invertebrate TIMPs are considered to represent the more ancestral TIMP form, from which possibly the rest of TIMPs originated. In order to understand thus the proximity of *d*TIMP to invertebrate TIMPs, and identify the rooting TIMPs and the evolutionary relationship that links the members of this group, sequence alignments, phylogenetic tree analysis and structural alignment were carried out for *d*TIMP against the homology models generated for invertebrate TIMPs.

Sequence alignment of *d*TIMP against invertebrate TIMPs yielded a homology of 23%. Interestingly, invertebrates "lack" the equivalent of the vertebrate TIMP C-terminal domain (Figure 4.10). In contrast however to other invertebrates, the two TIMP alleles from *hydra* are full-length proteins, consisting of both N and C-terminal domains.

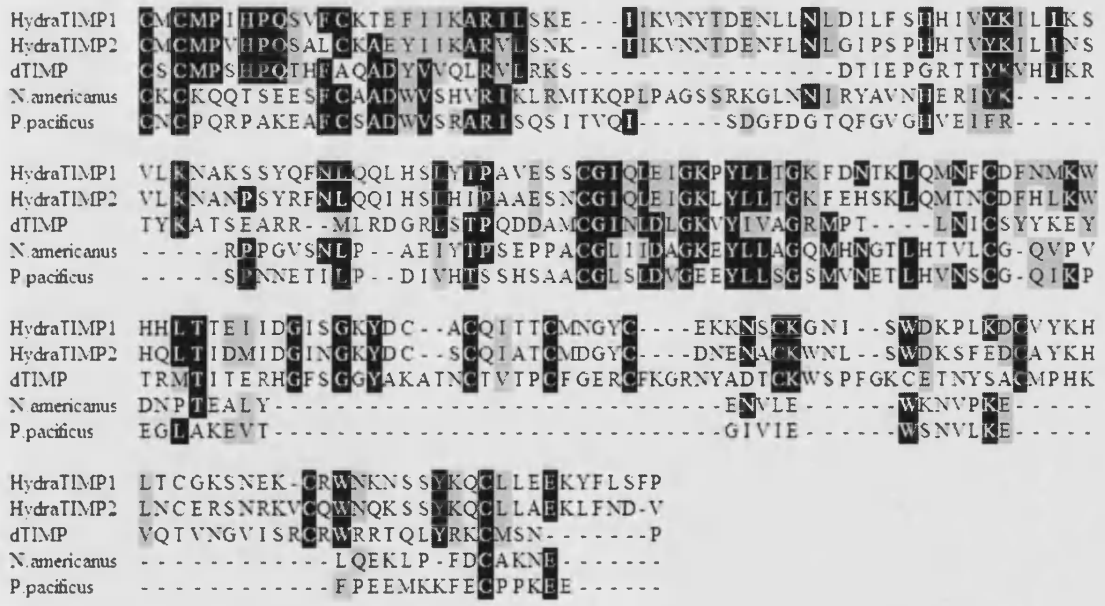


Figure 4.10: Sequence alignment of *d*TIMP against invertebrate TIMPs. Invertebrate TIMPs, with the exception of the two TIMP alleles from *hydra*, appear to 'lack' the equivalent to the vertebrate C-terminal domain.

To identify whether the lack of the C-terminal domain from invertebrate TIMPs is the sole reason for this low sequence homology, sequence alignment was carried out as previously, with one difference: C-terminal domains of *d*TIMP and the two *hydra* TIMPs were deleted and their N-terminal domains were aligned against the invertebrate TIMPs (Figure 4.11). This yielded a sequence homology of 30%, slightly only higher than that obtained previously. It is however possible to suggest that the C-terminal domain in *hydra*, *d*TIMP and vertebrate TIMPs is most probably an evolutionary acquisition.

Figure 4.11: Sequence alignment of N-*d*TIMP and N-*hydra*TIMP1 and N-*hydra*TIMP2 against invertebrate TIMPs. It is evident that even when the C-terminal is removed from *d*TIMP and the two *hydra* TIMPs, the sequence similarity with the invertebrate TIMPs is low.

Multiple sequence alignments of both sets of sequences (with and without the C-terminal domains in *d*TIMP and *hydra* TIMPs) indicated several differences of the sequences examined. In both cases most regions of the sequences are not conserved for *d*TIMP and invertebrate TIMPs, with the exception of the two *hydra* TIMPs. Due to the similarities observed for *d*TIMP and *hydra* TIMPs, the sequence homology was calculated for these sequences only, and was found to be around 49%. This is significantly higher than the homology percent obtained for the comparison of *d*TIMP against all invertebrate TIMPs. This finding indicates the proximity of *d*TIMP to *hydra* TIMPs rather than invertebrates

TIMPs, as a whole. This proximity of the three TIMPs would thus be expected to be evident from the phylogenetic tree analysis as well.

Both sets of sequences examined, (full-length sequences and same set of sequences, after removing the C-terminal domain from *d*TIMP and *hydra* TIMPs) yielded the same clusters in the phylogenetic tree (hence the phylogenetic tree of the second set of sequences is not included). The root of the tree is formed by *d*TIMP and *P.pacificus*. *Hydra* TIMP1 'branches off' from *d*TIMP, while *hydra* TIMP2 appears to occur later. *N.americanus* seems to 'branch off' from *P.pacificus* indicating the proximity of these two invertebrate TIMPs and their difference from *hydra* TIMPs and *d*TIMP (Figure 4.12). Hence, as expected from the sequence homology and multiple sequence alignment, the two *hydra* TIMPs are most proximal to *d*TIMP than the other representatives of invertebrate TIMPs.

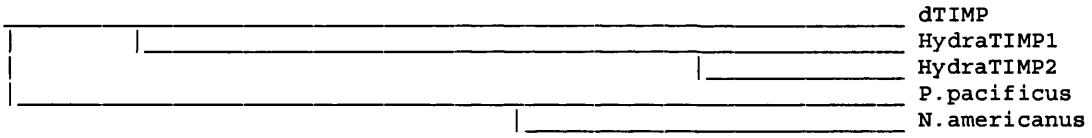


Figure 4.12: Phylogenetic tree analysis of *d*TIMP against invertebrate TIMPs. The root of the phylogenetic tree is generated by *d*TIMP and *P.pacificus* while the rest of the invertebrate TIMPs occupy later branches.

As no structures have been reported for invertebrate TIMPs, homology models were generated and aligned against the model generated for *d*TIMP. The results obtained from the alignment of the models are provided in Table 4.4.

Table 4.4: Structural alignment of *d*TIMP against invertebrate TIMPs.

Structures/models aligned	RMSD (Å)
*d*TIMP–*Hydra* TIMP1	1.4
*d*TIMP–*Hydra* TIMP2	1.5
*d*TIMP–*Necator americanus* TIMP	1.3
*d*TIMP–*Pristionchus pacificus* TIMP	1.4

RMSD is calculated for backbone (C^{α}) only using the program SPDBV (Guex and Peitsch, 1997).

From all homology models of invertebrates aligned against *d*TIMP, *N.americanus* aligned better against *d*TIMP, with an rmsd of 1.3Å, followed by *P.pacificus* with an rmsd of ~1.4Å. *Hydra* TIMP1 and TIMP2 -shown to be most proximal to *d*TIMP from the phylogenetic tree analysis and the sequence homology- yielded rmsd of 1.4Å and 1.5Å respectively. It is highly likely that the results obtained for the alignments of *d*TIMP against invertebrate TIMPs are not sufficiently accurate, as the latter group of TIMPs consists of only one domain, which probably aligns well against the N-terminal domain of *d*TIMP, as this domain is considered to be conserved and bare the conserved inhibitory activity of TIMPs. In contrast, the two *hydra* TIMPs, which are two-domain proteins probably have a higher rmsd against *d*TIMP due to differences in both N and C-terminal domains of the homology models.

4.5.3 dTIMP against insect TIMPs

Two putative TIMPs have been more recently identified in insects *Anopheles gambiae* and *Aedes aegypti*. The genome of *Anopheles gambiae,* the principal vector of malaria in Africa, was sequenced in 2002 (Holt *et al.*, 2002). On chromosome 2R, a locus (locus tag ENSANGG00000012010; primary accession number A0NG61) has been identified to be similar to tissue inhibitor of metalloproteinase from *Drosophila melanogaster* (*d*TIMP). This is a 216 amino acid sequence in length, similar to the usual TIMP length that ranges

from 180 to 210 amino acid residues. The conserved C-X-C motif at the N-terminal domain of mature TIMPs is also present in this protein, preceded by a leader sequence of 28 amino acid residues. Preliminary analysis on this sequence lead to the identification of an NTR domain, characteristic of TIMPs, and thus allowed its classification as TIMP1. Genome sequencing of *Aedes aegypti,* a tropical mosquito that is frequently used as a host for dengue fever and yellow fever viruses, (Nene *et al.*, 2007) lead to the identification of a 213 amino acid protein, with the characteristic C-x-C motif on the N-terminal domain. Preliminary analysis on this protein revealed the presence of an NTR domain, and was classified as a putative TIMP3. In order to identify the proximity of the two insect TIMPs to *d*TIMP, as well as to vertebrate and invertebrate TIMPs, an analysis of all TIMPs was carried out.

Sequence alignment of the three insect TIMPs, *A.g*TIMP, *A.a*TIMP and *d*TIMP and calculation of their sequence homology, reveals that they share an overall homology of 55% with several conserved regions, as seen from the output of boxshade (Figure 4.13). Hence, it is possible to say that although the classification of the two insect proteins as TIMPs is preliminary, the high sequence homology to *d*TIMP and the conserved areas as seen from the sequence alignment, confirm their common ancestry.

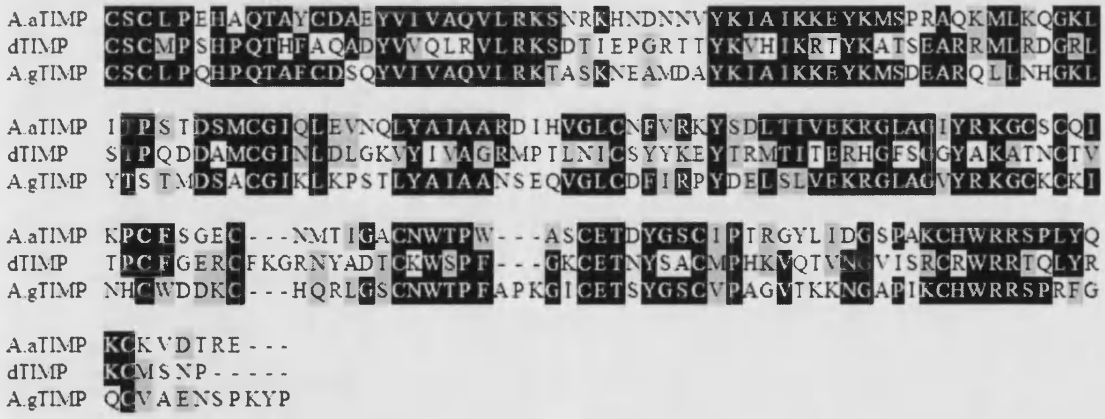


Figure 4.13: Boxshade output for the sequence alignment of the three insect TIMPs: *A.a*TIMP, *d*TIMP and *A.g*TIMP. Several regions appear to be conserved among the three insect TIMPs. Shown in black background are the conserved amino acid residues, shown

in grey are the equivalent amino acid residues and in white background are residues which are different in the three insect TIMPs.

Analysis of insect TIMPs against vertebrate TIMPs yielded a homology of 43%, while the homology against invertebrate TIMPs was only 28%, which is within the same range as the percentage homology obtained for full-length *d*TIMP against invertebrate TIMPs. It is important however to remember that invertebrate TIMPs are lacking the vertebrate C-terminal domain and are thus much smaller. In order to evaluate the proximity of TIMPs from different species, a phylogenetic tree analysis of all TIMPs used in this study was carried out (Figure 4.14).

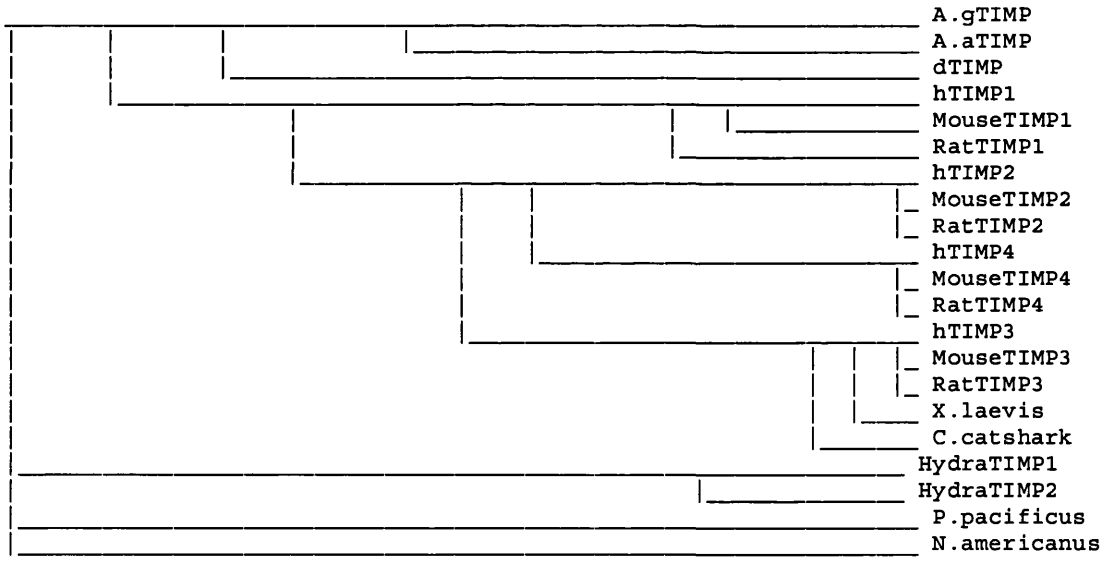


Figure 4.14: Phylogenetic tree analysis of insect TIMPs against vertebrate and invertebrate homologues. The root of the phylogenetic tree is formed by invertebrate TIMPs, from which the rest of TIMPs originated.

As seen in the phylogenetic tree analysis (Figure 4.14) the most ancestral TIMP forms are those present in invertebrates, and insects, as well as their human TIMP1 homologue. These are followed by the rest of the vertebrate TIMP homologues: of the four classes of

vertebrate TIMPs, TIMP2, TIMP3 and finally TIMP4. It is hence possible to suggest, based on these findings, that invertebrate TIMPs represent the most ancestral TIMP form, from which the insect and vertebrate TIMPs originated. Conservation of disulphide bonds and the characteristic C-x-C motif in the N-terminal domain of all TIMPs verifies their common ancestry and their evolutional proximity.

Phylogeny provides information on evolutionary history, showing how organisms are related (rather than providing evidence for evolution as such). All organisms examined in this study are eukaryotes and belong to the animal kingdom (metazoa). Most ancestral of all is *hydra* belonging to the phylum cnidaria. Nematodes are the next in line, followed by *d*TIMP and finally the vertebrates. *Hydra* TIMPs, seen as rooting sequences in the phylogenetic trees, provide valuable information on the evolutionary relationship of TIMPs.

A closer look at the amino acid sequence of TIMPs from different species reveals a detail of key significance. Invertebrate TIMPs, as one-domain molecules, contain a total of six cysteines, probably all involved in disulphide bonds. It is however not possible to extrapolate that to vertebrates TIMPs, which have an extra domain. Hence, it would be 'incorrect' to include invertebrate TIMPs in the following type of comparison, as due to the lack of one domain, all assumptions made would be inaccurate. Insect TIMPs present an interesting discrepancy: *A.a*TIMP and *A.g*TIMP contain a total of fourteen cysteines, by two more than those observed in *d*TIMP and vertebrate TIMPs. This is also the case for the two TIMP alleles present in *hydra*: a total of fourteen cysteines would account for seven disulphide bonds. As seen from Figure 4.14, *hydra* and insect TIMPs appear 'earlier' in the phylogenetic tree than some of their vertebrate homologues (with the exception of *h*TIMP1), possibly suggesting that that the total number of cysteines, conserved in more ancestral TIMPs, are fourteen and the conserved disulphide bonds are therefore seven. The reduction in the number of disulphide bonds observed in *d*TIMP and vertebrate TIMPs can be considered to be a result of evolution and the different needs of different species. Even if this is the case however, *d*TIMP differs from the rest TIMPs that have six disulphide bonds, in that the organisation of these bonds is not the same as

that seen in its vertebrate homologues. Based on these findings, it is possible to suggest that *d*TIMP is an 'evolutionary intermediate', between the most ancestral TIMP form and the vertebrate TIMPs which appeared later.

4.6 Conclusion and future work

The turnover of the extracellular matrix is a crucial process proven to control cellular behaviour. Normal physiological processes depend on the turnover of the components of the extracellular matrix (Nagase and Woessner, 1999; Lund *et al.*, 1999), which upon misregulation leads to several pathological conditions (see Chapter II – Introduction to angiogenesis) (Halpert *et al.*, 1996; Konttinen *et al.*, 1999; Nelson *et al.*, 2000). One of the major classes of enzymes, involved in the turnover of the extracellular matrix are the MMPs which are responsible for the degradation of the extracellular matrix. Regulation of MMP activity, which is crucial for the living organism, occurs either by transcriptional control of MMPs, their secretion in an inactive form (zymogens) or by their inhibition. Inhibitors of MMPs, the TIMPs, present a potent means for the design of inhibitors of MMPs in pathological conditions, where MMPs are overexpressed.

Research on TIMPs and their evolutionary relationship has suggested that of the four TIMP alleles (present in humans and most mammals) TIMP1 is most proximal to the ancestral TIMP form, from which the rest of the TIMP classes originated, probably *via* gene duplication (Brew *et al.*, 2000). In order to understand the evolutionary relationship that links TIMPs from different organisms, a representative of the more ancestral TIMP form was examined. *d*TIMP differs from vertebrate TIMPs in that it lacks one of the conserved disulphide bonds from the N-terminal domain, which is the catalytic domain of TIMPs, while it has an extra disulphide bond in the C-terminal domain, thought to assist in the orientation and binding of TIMPs to the MMPs they inhibit. It was not clear whether this difference represented an indication of ancestry. Bioinformatic and modeling analyses on *d*TIMP against vertebrate, invertebrate and two insect TIMPs provided a wealth of information regarding the conserved disulphide bonds, their possible organisation and the evolutionary relationship that links TIMPs from different species.

Sequence homology and phylogenetic tree analysis of vertebrate TIMPs against *d*TIMP, revealed that, *h*TIMP1and *h*TIMP3 are most proximal to the ancestral TIMP form, represented in that comparison by *d*TIMP. Interestingly this result is in agreement with two previous findings, one of which suggested that TIMP1 is most proximal to the ancestral TIMP form, from which the rest of the TIMP classes originated *via* gene duplication (Brew *et al.*, 2000), while the other suggested that *d*TIMP is most similar to mammalian TIMP3 (Wei *et al.*, 2003). As the structure of *d*TIMP has not yet been reported, it would not be possible to be conclusive on its proximity to either *h*TIMP1 or *h*TIMP3, at the structural level. Sequence homology of vertebrate TIMP2 and TIMP4 against *d*TIMP is around the same range, indicating that these two TIMP classes probably evolved in parallel, although the phylogenetic tree analysis places TIMP2 directly after TIMP1. Alignment of the different vertebrate TIMP classes against the homology model generated for *d*TIMP revealed only a few differences, as expressed by the low deviation of the rmsd calculated for the above structural alignments. It is important to note that all members of vertebrate TIMPs have conserved organisation of disulphide bonds, which differs from that observed in *d*TIMP.

Sequence homology and phylogenetic tree analysis of invertebrate TIMPs against *d*TIMP provided some interesting information. The low overall sequence homology, can be most probably attributed to the lack of one domain from the two representatives of invertebrate TIMPs, *P.pacificus* TIMP and *N.americanus* TIMP. Proximal to *d*TIMP however were shown to be the two TIMPs from *hydra*, as expressed by the sequence homology and their proximity in the phylogenetic tree analysis. The two TIMP alleles present in *hydra* yielded an rmsd of 1.4-1.5Å, similar to that obtained from the alignments of vertebrate TIMPs against *d*TIMP. Phylogenetic tree analysis revealed that *hydra* TIMPs form probably the root of TIMPs, from which the rest originated. Interestingly, the more ancestral, rooting *hydra* TIMPs, have more disulphide bonds than their vertebrate homologues, probably suggesting that the number and organisation of disulphide bonds in vertebrate TIMPs are an evolutionary acquisition.

Including insect TIMPs in the analysis provided some novel results. As seen from the primary sequence of the two insect TIMPs and the phylogenetic tree analysis, they both contain an extra disulphide bond from *d*TIMP and vertebrate TIMPs. If the disulphide bonds and their organisation are used as a basis for the analysis, it would be possible to suggest that the two TIMPs from *A.gambiae* and *A.aegypti* are more proximal to the ancestral TIMP form (e.g. *hydra*). The organisation of disulphide bonds in *d*TIMP would classify it as an intermediate, between the ancestral TIMPs (an extra disulphide bond, resulting in extra stabilisation in the C rather than the N-terminal domain of TIMPs) and the more recent vertebrate TIMPs (different organisation of disulphide bonds in the two domains). This hypothesis would explain the differences between *d*TIMP, *hydra* and insect TIMPs, as well as between *d*TIMP and vertebrate TIMPs.

As an intermediate, *d*TIMP could offer a wealth of information on the evolutionary relationship of TIMPs as well as on the significance and role of the conserved disulphide bonds, a characteristic TIMP feature. *d*TIMP has been shown to be a potent inhibitor of MMPs and TACE, indicating that the lack of one disulphide bond from the N-terminal domain does not reduce or abolish its enzymatic activity. Further work on more ancestral TIMP representatives, such as TIMPs from *A.gambiae* and *A.aegypti* as well as the two TIMP alleles from *hydra*, would indicate whether the different organisation of the disulphide bonds allow these molecules to still be active, even against the vertebrate MMPs.

Identification of more ancestral MMPs forms, such as those preset in insects and *hydra* would show whether there are any differences with their vertebrate homologues. It is possible that the higher number of disulphide bonds observed in more ancestral TIMP forms, were required for the inhibition of MMPs of these organisms. It is thus possible that also the MMPs in insects and *hydra* also have differences from those present in *Drosophila melanogaster* or vertebrates.

Structural studies on TIMPs from different species would provide further information on the significance of conserved features and how these could be used for the design of new

drugs, specific for the inhibition of MMPs. Structural studies on more ancestral TIMP forms would enhance the understanding of the structural features of these molecules and their significance. Biological experiments on ancestral TIMP forms and the MMPs they inhibit would provide interesting and novel information on the evolution of these molecules and their potential use in therapeutics. It would be interesting to examine whether invertebrate TIMPs, which are smaller in size than their vertebrate homologues, could be used for drug design against vertebrate MMPs.

Appendix I – Fermentation

Protein expression is most commonly carried out in 2L flasks, containing media and antibiotics of choice. Cell growth is monitored by measuring the optical density (O.D.) at 600nm and when it reaches 0.6 to 0.8, protein expression is induced by the addition of IPTG (isopropyl β-D-thiogalactoside). IPTG induces the T7 promoter, behind which the gene of interest is inserted, inducing in this way the expression of the protein of interest. Another commonly used method for protein expression is the use of auto-induction media, where expression is not induced by the addition of IPTG. In this case expression media contain a mixture of carbon sources, including lactose. Of all carbon sources glucose is usually depleted first. Lactose can then enter the cell and induce the expression of T7 polymerase, inducing in this way protein expression. This type of protein expression is advantageous in that it does not require constant monitoring. These two types of protein expression are most commonly tested in order to identify which produces a higher protein yield, preferably in the soluble phase. Optimisation of the conditions used (e.g. temperature, incubation time post-induction and addition of extra carbon sources) is usually required.

Both types of protein expression described above were tested for hTP expression. Induction of protein expression using IPTG proved to be advantageous over auto-induction method. None of the two methods however yielded a satisfactory amount of soluble protein. As a large amount of protein is required for biological experiments and crystallisation trials, overexpression of hTP was required. In order to achieve a higher protein yield a large scale protein expression experiment was set up. Large scale hTP expression was carried out in a fermentor, BioFlo 3000 (New Brunswick Scientific), where the environment is more 'controlled' than that of cultures in 2L flasks. As a new method, hTP expression using the fermentor required several rounds of optimisation until the protein yield was satisfactory.

Optimal conditions for hTP large-scale expression were 30°C, pH of 6.5 and agitation which was set in-loop with the dissolved oxygen (D.O.; see Chapter III, *3.9.2*). Antifoam was added to the cultures in µl quantities, as excessive addition of antifoam can have

adverse effects: inhibition of cell growth. Cultures were incubated for 3hrs post-induction and harvested through the harvest port of the fermentor (Figure 1; Figure 2).

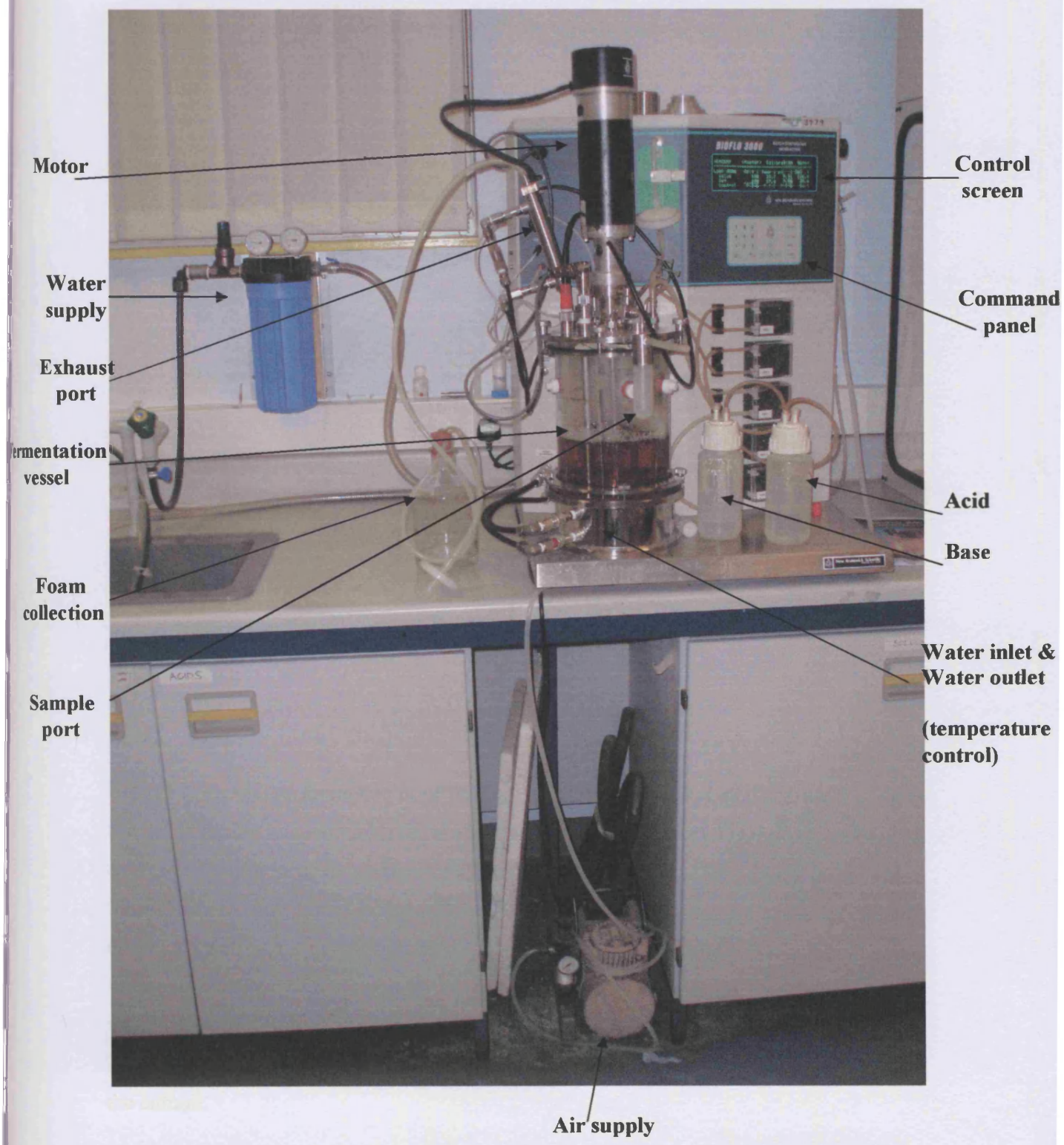


Figure 1: BioFlo 3000 (New Brunswick Scientific). This figure illustrates the fermentor apparatus.

The headplate organisation of the BioFlo 3000 (New Brunswick scientific) is illustrated below; Figure 2.

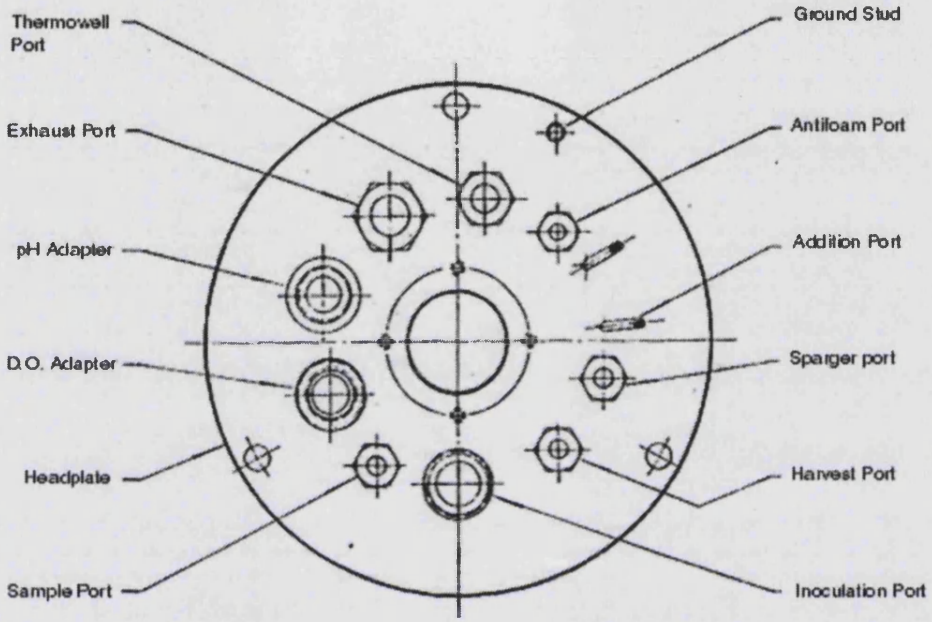


Figure 2: Headplate arrangement of the BioFlo 3000 fermentor. Labeled are the ports in the headplate for temperature (temperature probe), pH (pH probe), D.O. (D.O. probe) and antifoam (antifoam probe). Any foam produced during bacterial growth is discarded through the exhaust port. The inoculation port is used for the addition of the inoculating culture as well as for antibiotics. The sample port (which essentially is a valve allowing for small amounts of sample to be collected) is used for monitoring the culture growth while harvesting is done through the harvest port. Two additional ports are also available. These are used for the addition of acid and base to the vessel so as to control the pH of the culture.

Large-scale hTP expression (fermentor) resulted in a much higher yield that than previously obtained from cultures in 2L flasks. A 3.5L fermentor culture yielded 700µl of pure hTP (concentration of 9mg/ml). This is much higher than that previously obtained using a standard protocol for 3.5L culture in culture flasks. This standardised method usually yielded 200µl of 6-7mg/ml concentration. It is thus evident that large-scale expression in a controlled environment proved to be advantageous for hTP expression.

Fermentor preparation

Preparation of the fermentor for use contains several steps which should be followed carefully. All parts of the fermentor need to be thoroughly washed with soap water and rinsed with de-ionised water. The use of ethanol in the vessel is not recommended, as the rubber rings in the vessel can be damaged. The vessel is assembled and all the tubing is attached to the respective ports. The tubing used is made of silicon, while plastic strings are used to tighten the connection of the tubes to their respective ports. The pH and D.O probes require calibration prior to every use. The pH probe, stored in 3M KCl, can be calibrated using the solutions typically used for pH meter calibration. The D.O. probe requires frequent monitoring as it is very sensitive. To ensure correct use it is essential to check that it has sufficient amount of electrolyte, which aids in the sensitivity of the measurements for the dissolved oxygen. Once the pH probe is calibrated, the vessel (containing the media and attached to all the tubing) can be autoclaved at 121°C, 15psig for 45min. Following autoclave the vessel is allowed to cool down and then attached to the main fermentor unit, through which the vessel is controlled.

The vessel is connected to the main unit and to a supply for water and air (Figure 1). Water is necessary for the calibration of temperature and air is essential for culture growth. Vessel control is done via the screen of the main fermentor unit (Figure 1). The main screen allows access to three 'sub-screens': 1) a calibration screen - used for calibration of the temperature, pH and D.O. probes, 2) a master screen - which allows the monitoring of the vessel during the culture and provides readings as well as the user-set

values and 3) a gases screen - through which control of vessel aeration is done and additional air can be supplied to the vessel.

Upon connection of the vessel to the main fermentor unit, the first step is to prime the temperature to the required value for culture. To ensure the temperature is measured correctly a small amount of glycerol is added to the thermowell before the temperature probe is connected. The temperature probe is allowed to equilibrate and once the desired temperature is reached the D.O. probe can be calibrated. For calibration of the D.O. probe two extremes are required: a minimum and a maximum. Minimum D.O. readings are obtained by disconnecting the D.O. probe from the main fermentor unit and setting the D.O. reading to zero. The vessel is allowed to equilibrate and the D.O. probe is again connected to the vessel. To achieve the maximum readings, the agitation is set to the maximum value that will be used during the experiment and the D.O. probe is again allowed to equilibrate. In order to ensure the correct amount of air is going into the vessel, the D.O. can be set in-loop with agitation, which is a more automatic way of controlling vessel aeration.

The additional ports are connected to acid and base solutions (Figure 1), if pH control is required for the experiment. Most commonly used acid solution is acetic acid (hydrochloric acid is not recommended) while the base solution most commonly used is ammonium hydroxide. All solutions should be autoclaved prior to use in order to minimise the chance of contamination.

Once the vessel has been correctly calibrated and equilibrated, the media is inoculated through the inoculation port. The same port is used for the addition of antibiotics. Monitoring of culture growth is facilitated by the sample port which allows for samples to be taken during the experiment, without disrupting the rest of the culture. The continuous use of agitation ensures that the sample used for monitoring culture growth is homogeneous. When cell density reaches 0.6-0.8 at $O.D._{600}$ nm the culture can be induced via the inoculation port. Upon inoculation a sample is collected, namely T_0, which can be later used for SDS-PAGE analysis against a sample of the harvested culture (assuming

maximum growth has been reached and the protein has expressed after the addition of IPTG).

Typical of culture growth (Figure 3) is the increase in the absorbance readings at $O.D._{600}nm$ over time which is connected to a decrease in the oxygen available in the vessel (Figure 3). To ensure that oxygen depletion in the vessel does not become an inhibitory factor for cell growth, the agitation can be set in-loop with the dissolved oxygen. In this way agitation increases when D.O. is reduced, allowing in this way for more aeration, which will provide the oxygen required for cell growth.

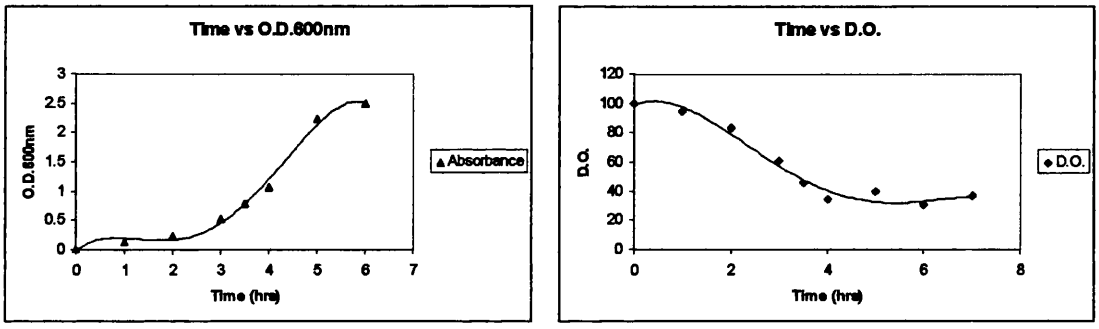


Figure 3: Plots illustrating bacterial growth during fermentation: the relationship between time and absorbance at $O.D_{600}$ nm is indicative of the growth, which results at the reduction of D.O. remaining in the vessel at the same time.

It is likely that during bacterial growth, foam will be formed in the vessel. In this case the exhaust port is used. If the foaming is excessive, antifoam solution can be added to the vessel *via* the inoculation port. It is essential to add as little as possible, since excessive use of antifoam can lead to inhibitory effects on culture growth. Harvesting of the culture occurs through the harvesting port.

Upon completion of the experiment the main unit has to be shut down and the water and air supply have to be disconnected. Vessel and tubing have to be washed with soap water and

rinsed with de-ionised water. The probes need to be carefully removed from the vessel assembly and rinsed with de-ionised water.

Optimisation steps for protein expression using the fermentor are required, as the use of a controlled environment is not always favourable under the same conditions, for the expression of all proteins. Once the fermentation conditions are optimised for each type of protein expression required, the yield obtained can be much higher. This allows the production of the protein of interest in a more efficient way, which is as user friendly as the commonly used types of protein expression. hTP is a characteristic example of the advantageous method of large-scale expression, as opposed to the standard culture method (2L flasks; see Chapter III, *3.9.2*).

Appendix II – Enzyme kinetics

Enzyme kinetics is the study of the rates of chemical reactions catalysed by enzymes. Studying the kinetics of a particular enzyme provides insights into several interesting areas, such as: the mechanism of action, ways to control the enzymatic activity and possible ways to inhibit the enzymatic activity. Understanding the information obtained from enzyme kinetics allows the understanding of the enzyme, its significance and role.

Rate of an enzymatic reaction

The rate of an enzymatic reaction (as in a chemical reaction) is expressed in terms of the change in the concentration of a product or reactant in a given time interval. This is more clearly illustrated by:

$$A + B \longrightarrow P$$

where A and B are the substrates in the reaction and P is the product. In this type of reaction, the rate will be expressed as:

$$\text{Rate} = -\frac{\Delta[A]}{\Delta t} = -\frac{\Delta[B]}{\Delta t} = \frac{\Delta[P]}{\Delta t}$$

where Δ is the difference; ΔA is the difference in concentration of substrate A, over a period of time, expressed as Δt. The rate of the reaction has a negative sign for the reactants (as they get consumed) and a positive sign for the product (as its production is increasing over time).

The rate of reaction (at a given time) is proportional to the product of the concentration of the reactants raised to the appropriate powers. More specifically,

$$\text{Rate} = k\,[A]^f\,[B]^g$$

Where K is the rate constant and where ^{f} and ^{g} are the exponents which must be determined experimentally.

The overall rate of a reaction is given by the sum of all the exponents. For example, a reaction where a reactant is directly converted to product, is a first order reaction. An example of a first order reaction is given below:

$$A \longrightarrow P$$

$$\text{Rate} = k\,[A]^1$$

First order reaction

When two reactants are required to form the product, then the rate of reaction for each of the reactants will still be one, but the reaction will overall be a second rate reaction. An example is given below:

$$A + B \longrightarrow C + D$$

$$\text{Rate} = k\,[A]^1[B]^1$$

Second order reaction

Several enzymes require a substrate in order to be active, the binding of which might trigger some changes in the conformation of the enzyme. Two models are mainly used in order to describe the enzyme-substrate binding: the lock-and-key model and the induced-fit model.

Lock-and-key model

In this type of model, a high degree of similarity is assumed to exist between the substrate (**S**) and the geometry of the active site cleft of the enzyme (**E**). Hence, the substrate binds the active site of the enzyme and 'locks' there upon formation of the enzyme substrate complex (**ES**), as illustrated below (Figure 1). There is however a disadvantage in this type of model: it does not account for any flexibility, a property which is known to exist in biomolecules.

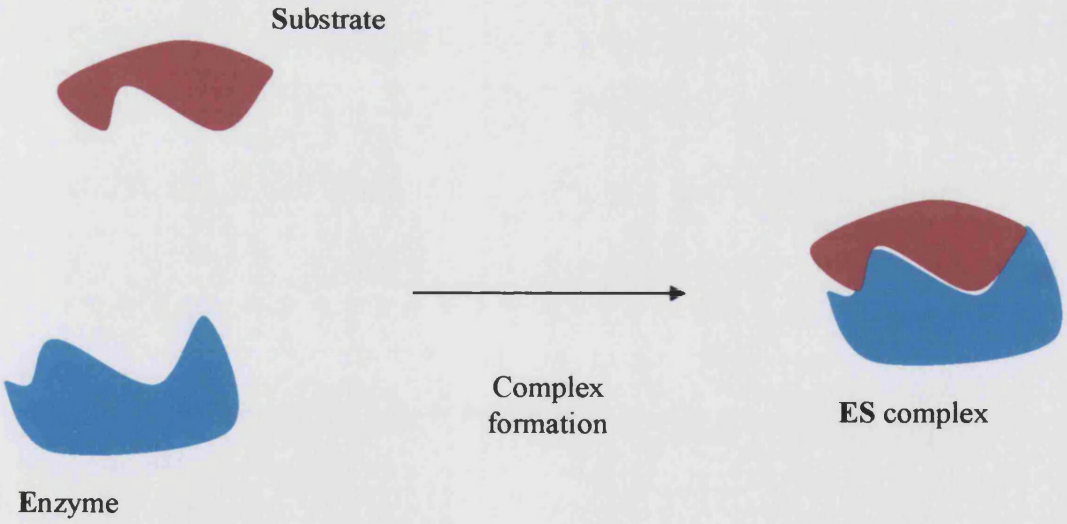


Figure 1: Lock-and-key model for enzyme-substrate complex formation. The active site of the enzyme is considered to have a high degree of similarity with the substrate. Upon binding, the substrate 'locks' in the active site of the enzyme and the reaction can occur.

Induced-fit model

Induced-fit model differs from lock-and-key in that it takes into account the conformational flexibility of biological macromolecules and does not assume that the geometry of the active site is highly similar to the geometry of the substrate. In this type of model, the binding of substrate is considered to induce a conformational change in the active site of the enzyme, which results in a complementary fit (Figure 2).

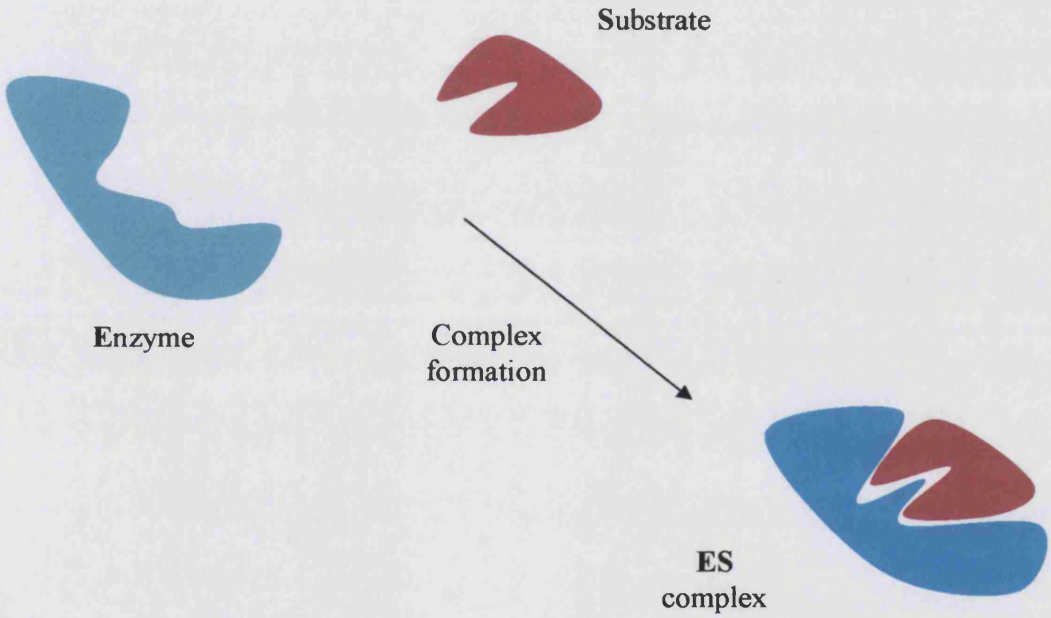


Figure 2: Induced-fit model for enzyme-substrate complex formation. The active site of the enzyme does not have similar geometry to the substrate. Binding of substrate induces conformational changes in the active site of the enzyme and its geometry now becomes complementary to that of the substrate.

Michaelis-Menten model

This model is the basic model used for the study of enzyme kinetics, especially for enzyme catalysed reactions. The main feature of the Michaelis-Menten model is the formation of an enzyme – substrate complex, as an intermediate of the reaction, which is relatively low but remains unchanged over the course of the reaction. Enzyme (E) and substrate (S) form a complex (ES) and although product (P) is formed during the course of the reaction, a low concentration of the complex (ES) always remains.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

To measure the rate of reaction (V_0) in this type of model, varying substrate concentrations ([S]) are required. In First-order kinetics, the rate of the reaction depends on the available concentration of substrate [S]. When the active site of all enzyme molecules are saturated with substrate, Zero-order kinetics are observed, where the rate of the reaction does not depend on concentration of [S] anymore (Figure 3).

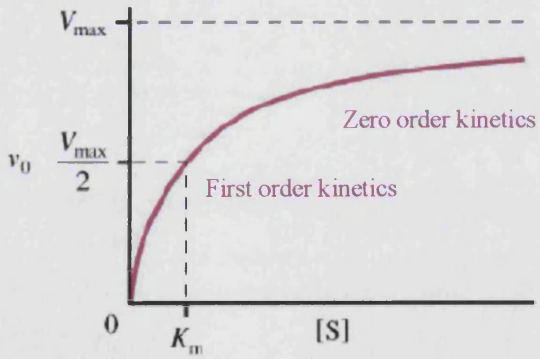


Figure 3: This figure illustrates the rate of an enzymatic reaction and its dependence on the available concentration of substrate. In the beginning of the reaction, when the active site of all enzyme molecules is not saturated, the rate of the reaction depends solely on

the available substrate concentration. When the enzyme becomes saturated with substrate, the available concentration of the latter does not affect the rate of the reaction.

According to the Michaelis-Menten model for enzyme catalysed reactions, an enzyme-substrate complex is always present during the course of the reaction. The rate of formation of the enzyme-substrate complex, is given by the following equation:

$$\textbf{Rate of formation} = \frac{\Delta[ES]}{\Delta t} = k_1\,[E]\,[S] \qquad (1)$$

The formed ES complex breaks down to free enzyme [E] and to product [P], while a part of the ES complex still remains present. The rate of breakdown of the ES complex, which leads to the formation of the product, is given by the following equation (The negative sign signifies the breakdown of the ES complex):

$$\textbf{Rate of formation} = -\frac{\Delta[ES]}{\Delta t} = k_{-1}\,[ES] + k_2\,[ES] \qquad (2)$$

A steady-state is reached when the rate of formation of the ES complex equals the rate of its breakdown. At this point,

$$\frac{\Delta[ES]}{\Delta t} = -\frac{\Delta[ES]}{\Delta t} \qquad (3)$$

and subsequently, $k_1\,[E]\,[S] = k_{-1}\,[ES] + k_2\,[ES]$ (4)

In equation (4) the enzyme and the substrate concentrations are known, as they are defined by the experimenter. The concentration of the ES complex as well as the enzyme which is still reacting are however not known. Hence, in order to solve this equation it is essential to know the concentration of all species in the reaction. From the initial enzyme concentration used for the reaction $[E]_T$ a part is present as free enzyme while some is involved in the formation of the ES complex. Hence the free enzyme, which still reacts will be given by:

$$[E] = [E]_T - [ES] \quad (5)$$

substituting (5) in (4)

$$k_1([E]_T - [ES])\,[S] = k_{-1}\,[ES] + k_2\,[ES] \quad (6)$$

collecting all the individual rate constants in (6),leads to (7):

$$\frac{[E]_T[S] - [ES]\,[S]}{[ES]} = \frac{k_{-1} + k_2}{k_1} = K_m \quad (7)$$

Solving (7) for [ES]:

$$[ES] = \frac{[E]_T\,[S]}{K_m + [S]} \quad (8)$$

In the Michaelis – Menten model, the initial rate (V_0) of product formation depends only on the rate of the breakdown of the [ES], and hence

$$V_0 = k_2\,[ES] \qquad (9)$$

substituting for [ES] in (**9**):

$$V_0 = \frac{k_2\,[E]_T\,[S]}{K_M + [S]} \qquad (10)$$

When the substrate concentration is too high all the enzyme is saturated and

$$[ES] = [E]_T$$

At this point, the reaction proceeds at its maximum velocity and

$$V_0 = V_{max} = k_2[E]_T$$

The original concentration of the enzyme is however constant. Hence

$$V_{max} = \text{constant}$$

Substituting the V_{max} expression in (**10**)

$$V_0 = \frac{V_{max}\,[S]}{K_m + [S]}$$

In general, the rate of reaction V in the Michaelis – Menten model, is given by the equation:

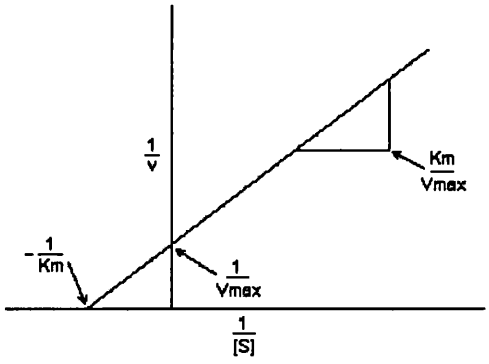
$$V = \frac{V_{max}[S]}{K_m + [S]}$$

Michaelis – Menten equation

By taking a reciprocal on both sides of the Michaelis – Menten equation results in a straight line which allows the calculation of K_m and V_{max} with more accuracy. This double reciprocal of the Michaelis – Menten equation is known as Lineweaver – Burk, and expressed as:

$$\frac{1}{V} = \frac{K_m + [S]}{V_{max}[S]} = \frac{K_m}{V_{max}}\frac{1}{[S]} + \frac{1}{V_{max}}$$

an example of the Lineweaver – Burk plot is given below:



From this plot K_m and V_{max} can be calculated with more accuracy. The x-axis represents the reciprocal of substrate concentration while the y-axis represents the reciprocal of

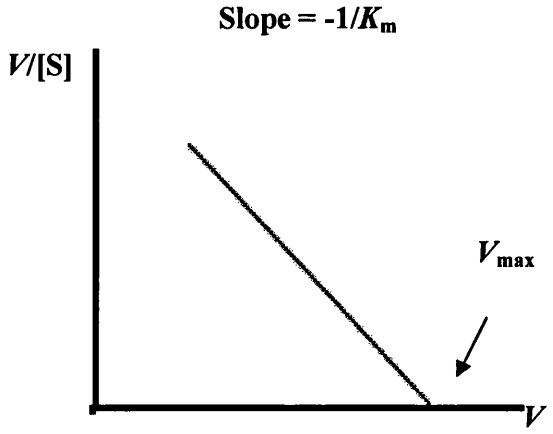
velocity. K_m is calculated from the x-axis intercept while V_{max} is calculated from the y-axis intercept. Hence the units of K_m are the same as those of the substrate and the units of V_{max} are expressed as concentration per time. It is essential to mention that the lower the value of K_m the higher the affinity of the enzyme for the substrate.

Apart from the Lineweaver – Burk plot for calculating K_m and $V_{max,}$ other methods are also available. These include the Eadie – Hofstee, the Hanes - Woolf and the Direct Linear methods.

(All methods described below produce linear plots and can be used for the accurate calculation of K_m and $V_{max.}$)

Eadie – Hofstee

In Eadie – Hofstee method the velocity is plotted on the x-axis while the y-axis represents velocity of substrate concentration.



The equation describing this method is the following:

$$\frac{V}{[S]} = \frac{V_{max}}{K_m} - \frac{V}{K_m}$$

The gradient is used for the calculation of K_m:

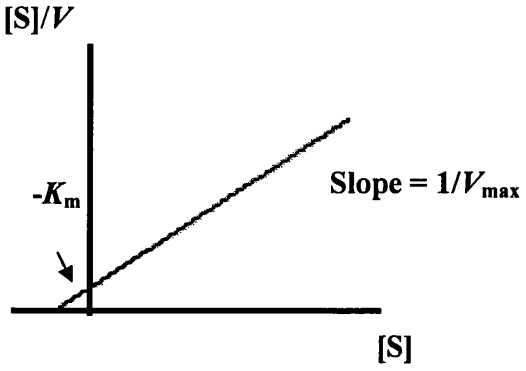
$$\textbf{Gradient} = -\frac{1}{K_m}$$

While V_{max} is calculated from the x-axis intercept.

$$\textbf{x-intercept} = V_{max}$$

Hanes – Woolf

In Hanes – Woolf method the substrate concentration is plotted on the x-axis while the y-axis represents the substrate concentration over velocity.



The equation describing this method is:

$$\frac{[S]}{V} = \frac{[S]}{V_{max}} + \frac{K_m}{V_{max}}$$

V_{max} is calculated from the gradient,

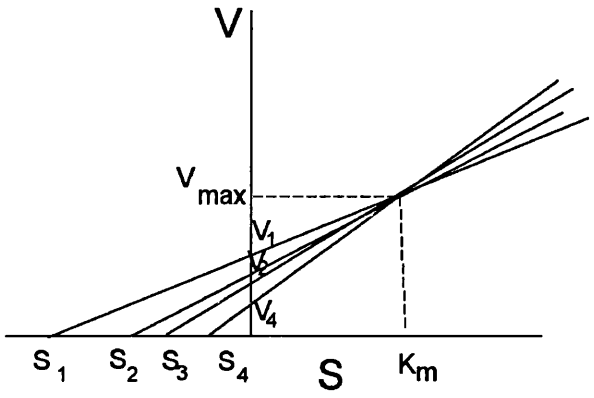
$$\textbf{Gradient} = -\frac{1}{V_{max}}$$

and K_m is calculated from the x-axis intercept,

$$\textbf{x-intercept} = -K_m$$

Direct linear

The direct linear method is considered to be one of the more accurate methods for the calculation of the kinetic parameters. In this type of method, lines are drawn for each of the substrate concentrations which are tested experimentally (e.g. S1, S2, S3, etc). Where these lines intersect on the x- and y-axis, the coordinates give K_m and V_{max} respectively. Due to the averaging of the data, direct linear is considered as a very precise method for the calculation of K_m and V_{max}. An example of the direct linear plot is given below:



The x-axis represents the substrate concentration and the y-axis represents the velocity of the reaction. The averaging of the data in this type of plot allows for a more precise calculation of the kinetic parameters K_m and V_{max}.

Once the values for K_m and V_{max} have been determined, it is essential to know how efficient the enzyme is in catalysing that particular reaction. This can be determined by the turnover number, which can also be calculated experimentally. The turnover number is the number of moles of substrate that react to form product per mole of enzyme per unit time. The units of the turnover number are sec^{-1}.

$$\frac{V_{max}}{E_T} = \textbf{Turnover number} = k_{cat}$$

This is indicative of the efficacy of the enzyme and can be used to illustrate how active the enzyme is as well as whether it is a good catalyst of a particular substrate.

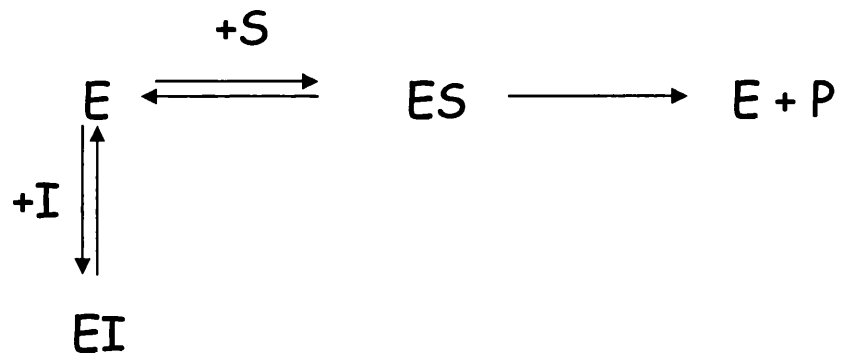
A simple enzymatic reaction can be described by the values obtained for K_m, V_{max} and k_{cat}. For the study of enzyme catalysed reactions where inhibitors are also present, more information is required for the description of the reaction.

In such cases the primary enquiry is the nature of the inhibitor and the type of inhibition. Once the type of inhibition is determined, the effect of the inhibition on K_m and V_{max} has to be determined.

Inhibitors can be generally described as substances that interfere with the activity of the enzyme, either reducing or abolishing it. In both cases the overall rate of the reaction is reduced and the effect of each type of inhibition can be determined. There are two main types of inhibitors: *Reversible* inhibitors bind the enzyme and are later released, leaving the enzyme in its original condition. *Irreversible* inhibitors react with the enzyme and the product is an enzymatically inactive protein. There are two major classes of reversible inhibitors: **Competitive inhibitors** and **Noncompetitive inhibitors.**

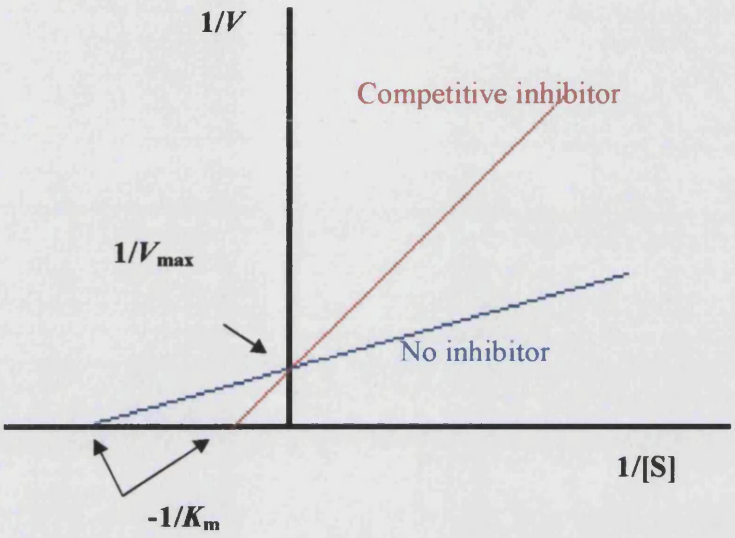
Competitive inhibition

In competitive inhibition, the inhibitor is very similar in structure with the substrate. It binds thus to the active site of the enzyme and blocks subsequent binding of the substrate. An example is given below:



It is evident that once the inhibitor binds to the active site of the enzyme, the substrate cannot bind and hence the reaction is inhibited.

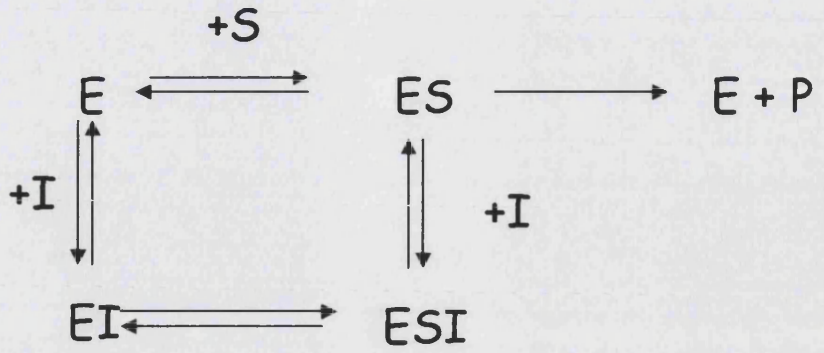
The Lineweaver – Burk plot of competitive inhibition compared to the plot of the same reaction, in the absence of inhibitor, is shown below:



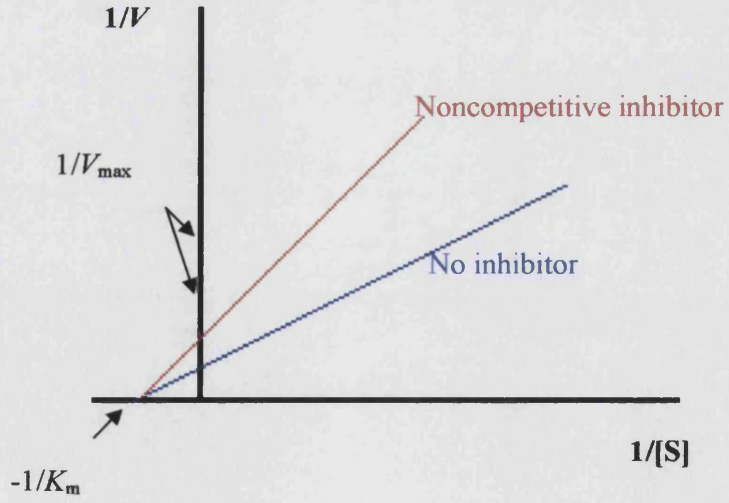
The Lineweaver – Burk plot for the enzyme catalysed reaction (without inhibitor) is illustrated in blue, while the reaction where inhibitor is also present is illustrated in red. From the plot it is evident that competitive inhibition yields a higher K_m value (as the enzyme's affinity for the substrate is lower since the inhibitor is competing with the enzyme for the active site occupation) while V_{max} is unaffected. A way to overcome this type of inhibition is to increase the concentration of substrate, as this would allow for more substrate to occupy the enzyme and drive the reaction.

Noncompetitive inhibition

In noncompetitive inhibition, the inhibitor binds at sites other than the active site of the enzyme. In this way, the substrate can still bind and the reaction can still proceed, although it will not be as effective. An example of non-competitive inhibition is given below:



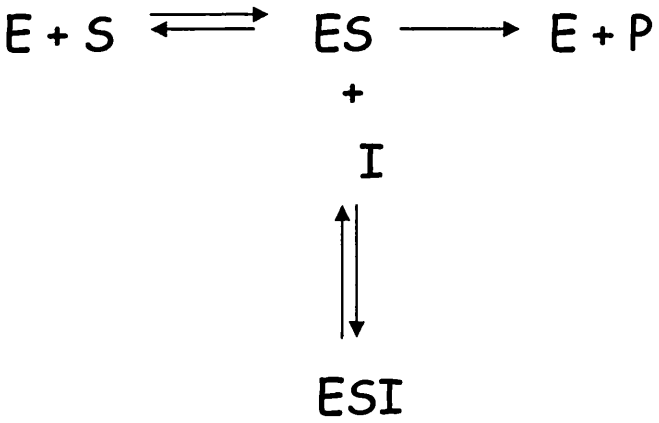
Binding of the inhibitor does not block the binding of substrate. A complex of enzyme-substrate and inhibitor is formed and the reaction can proceed. This reaction however will not be very efficient as the presence of the inhibitor will reduce the reaction rate. The Lineweaver – Burk plot of noncompetitive inhibition compared to the plot of the same reaction, in the absence of inhibitor, is shown below:



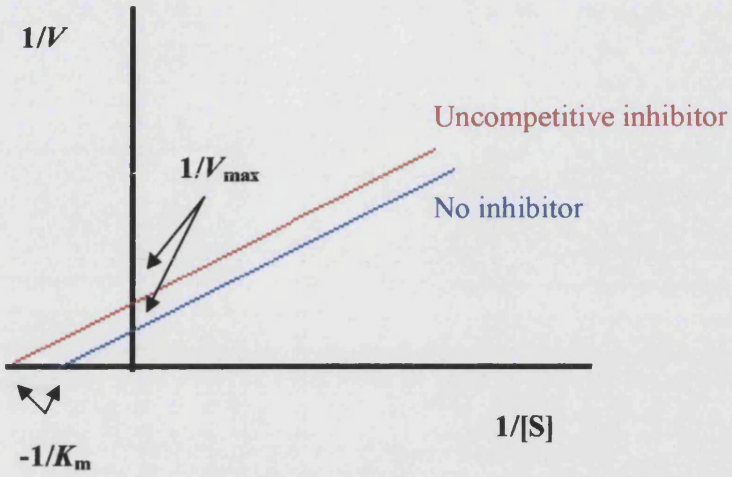
The Lineweaver – Burk plot for the enzyme catalysed reaction (without inhibitor) is illustrated in blue, while the reaction where noncompetitive inhibitor is also present, is

illustrated in red. From the plot it is evident that noncompetitive inhibition does not affect the K_m (as the inhibitor does not compete with the substrate for occupation of the active site of the enzyme) while V_{max} is reduced as the rate of the reaction is lower. As the inhibitor and the substrate are not 'competing' for occupation of the active, increasing the substrate concentration will not have any effect on the reaction.

Apart from the two most common types of inhibition, other types of inhibition also exist. These include **uncompetitive inhibition, mixed type inhibition** and **suicide inhibition.** In **uncompetitive inhibition** the inhibitor binds to the enzyme only after the appropriate substrate has bound. In this way the inhibitor inactivates the enzyme only after the formation of the ES complex. This type of inhibition can be described as:



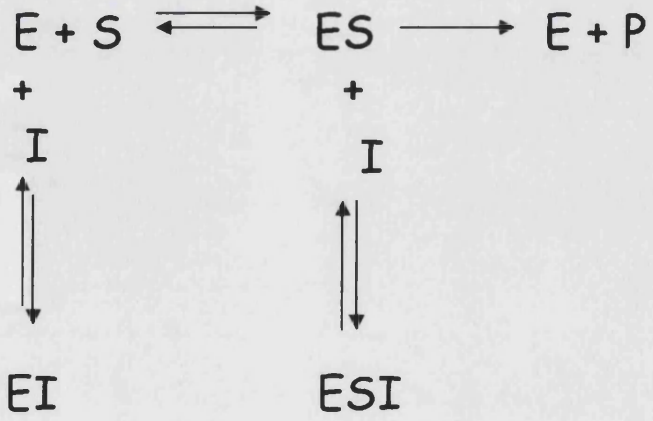
Uncompetitive inhibitors do not bind to free enzyme molecules; only after the binding of the substrate to the active site of the enzyme, and the formation of the enzyme-substrate complex, will the uncompetitive inhibitor bind and inhibit the reaction. The Lineweaver – Burk plot for this type of inhibition is shown below:



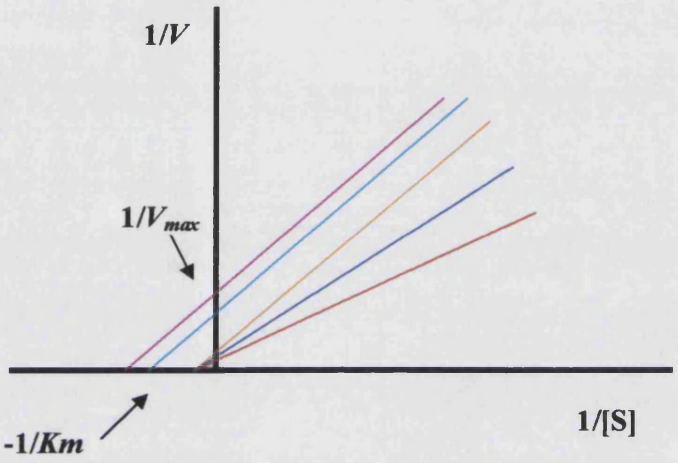
The Lineweaver – Burk plot for the enzyme catalysed reaction (without inhibitor) is illustrated in blue, while the reaction where uncompetitive inhibitor is present, is illustrated in red. From the plot it is evident that uncompetitive inhibition affects both K_m and V_{max} by reducing them. Increasing the substrate concentration does not affect the reaction in uncompetitive inhibition.

Mixed type inhibition

In this type of inhibition, the inhibitor can bind both to the free enzyme to give an enzyme-inhibitor complex, and to the enzyme-substrate complex to give an unreactive enzyme-substrate-inhibitor complex As shown below, both inhibitor-binding reactions are dead-end reactions and therefore reach equilibria.



Mixed type inhibitors can bind both the free enzyme and to enzyme-substrate complexes, inhibiting in this way the progression of the reaction. The Lineweaver – Burk plot for this type of inhibition is shown below:



The Lineweaver – Burk plot for the enzyme catalysed reaction (without inhibitor) is illustrated in blue, while the reactions where mixed type inhibitor is present, are illustrated in red, green, cyan and purple. From the plot it is evident that mixed type inhibition reduces the V_{max}, while K_m might increase or decrease.

Suicide inhibitors

Suicide inhibitors bind the enzyme and 'kill' it (inactivate it) while the inhibitor also 'dies' in the process. The reaction is irreversible: the enzyme and inhibitor are not released back into solution. These inhibitors are also known as **mechanism-based inhibitors**. Such inhibition is exploited in drug design and development (inhibitors targeting specifically an enzyme). The only way to understand this type of inhibition is by understanding the mechanism of enzymatic reaction and understanding how the enzyme becomes 'stuck' after inhibitor binding.

A table summarising the different types of inhibition and the effect they have on V_{max} and K_m is given below (Table 1).

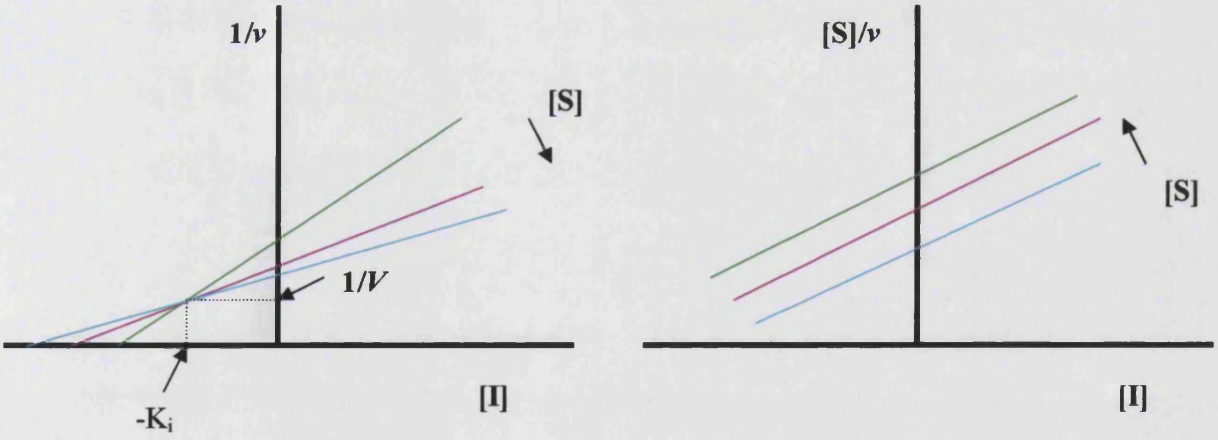
Table 1: Type of inhibitors and effect on the reaction

Type of inhibitor	Effect	Comments
Competitive	Increases K_m but does not affect V_{max}	Binds specifically at the active site, where it competes with the substrate for binding.
Noncompetitive	Reduces V_{max} but does not affect K_m	Binds E or ES complex other than at the catalytic site. Substrate binding unaltered, but ESI complex cannot form products. Inhibition cannot be reversed by substrate.
Uncompetitive	Reduces both K_m and V_{max}	Binds only to ES complexes at locations other than the catalytic site. Substrate binding modifies enzyme structure, making inhibitor- binding site available. Inhibition cannot be reversed by substrate.
Mixed	Affects both K_m and V_{max}	V_{max} is reduced K_m and may increase or decrease.

Inhibition efficiency

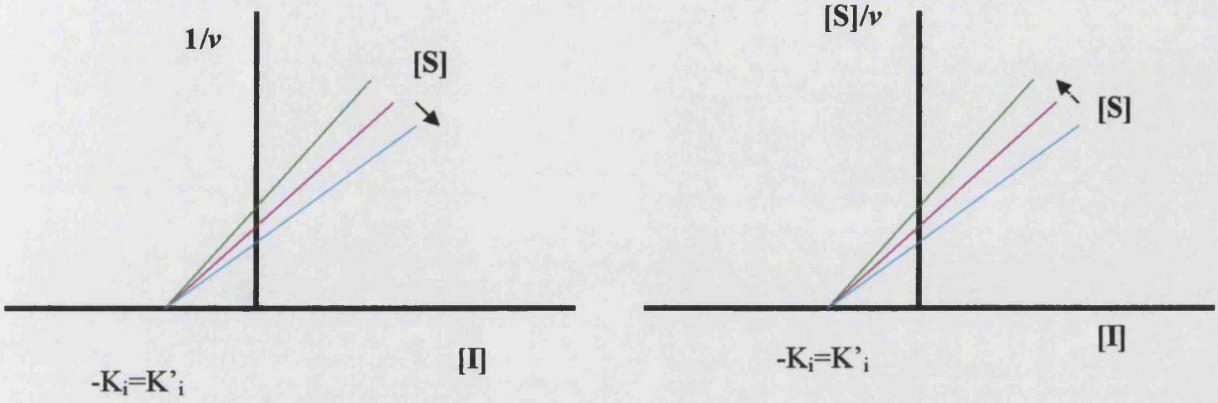
To 'measure' how efficient an inhibitor is, a constant can be used to describe its effect on the enzyme catalysed reaction. Inhibitor constant k_i is the equilibrium constant of the reversible combination of the enzyme with an inhibitor. k_i values can either be calculated or derived from Dixon plots. Dixon plots are plots of the inhibitor concentration versus the reciprocal of velocity. Each type of inhibition produces a different type of plot. Hence it is an accurate and easy way to determine both the type of inhibition as well as the efficiency of the inhibitor present. These plots however cannot be used to derive any information for the K_m and V_{max} parameters. Examples of Dixon plots for different types of inhibition are given below ([S] – substrate concentration; [I] – inhibitor concentration):

Competitive inhibition



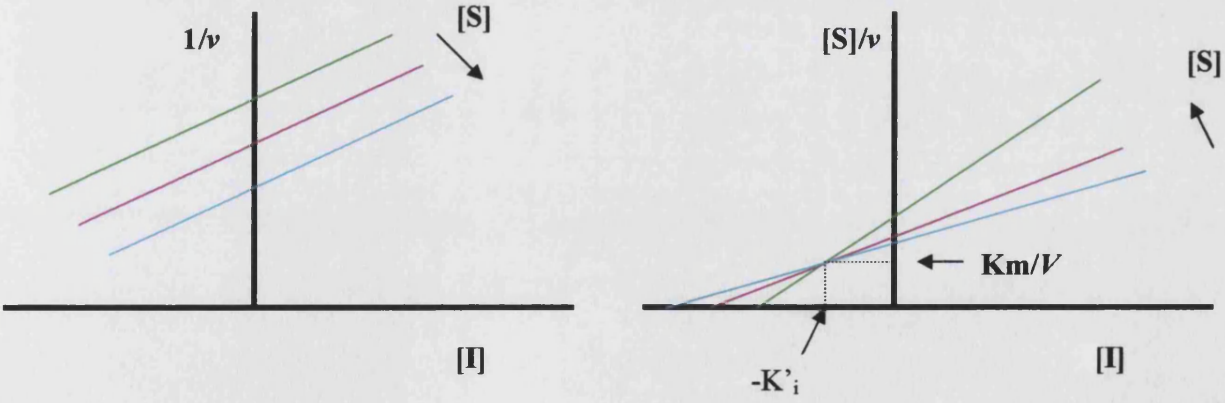
In competitive inhibition, when 1/v is plotted against [I], the substrate concentrations tested 'meet' above the x-axis. This plot allows the calculation of the Ki value for the inhibitor directly. In contrast, a plot of [S]/v against [I] (in competitive inhibition) produces parallel lines.

Non-competitive inhibition



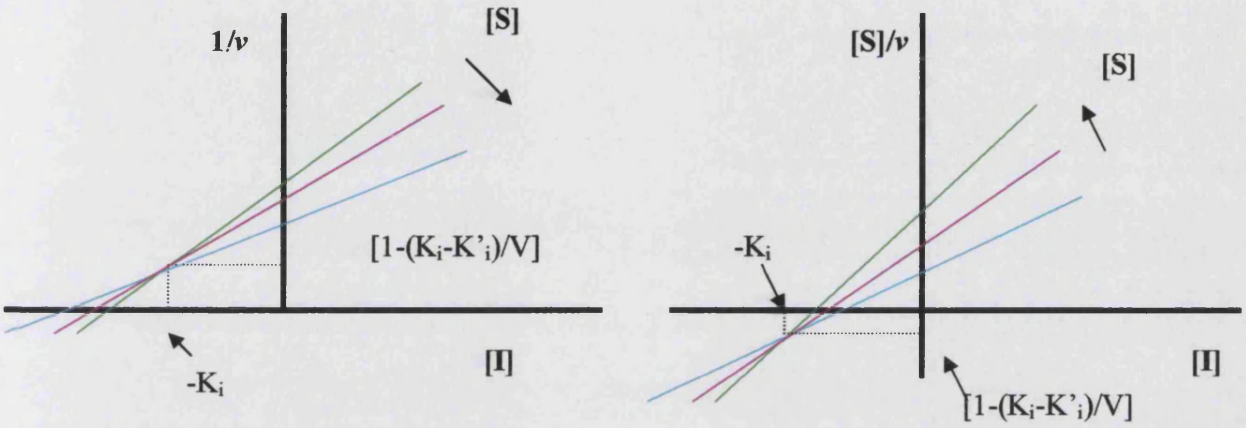
In non-competitive inhibition, when 1/v is plotted against [I], the substrate concentrations tested 'meet' on the x-axis, where $-K_i=K'_i$. In this type of inhibition, a plot of [S]/v against [I] is exactly the same as the one produced for 1/v against [I].

Uncompetitive Inhibition



In un-competitive inhibition, when 1/v is plotted against [I], the substrate concentrations tested produce parallel lines. In contrast, the plot of [S]/v against [I] produces lines that 'meet' above the x-axis and are used for the calculation of the K_i.

Mixed-type inhibition



In mixed-type inhibition, when 1/v is plotted against [I], the substrate concentrations tested meet above the x-axis, in contrast to the plot of [S]/v against [I], where the lines meet below the x-axis. The K_i value can be calculated from both plots, using the formula $[1-(K_i-K'_i)/V]$.

It is thus evident that Dixon plots are an efficient and convenient way for determining both the type of inhibition as well as the K_i value for a given inhibitor. These plots have been used for the determination of the type of inhibitor 5IUR is as well as for the calculation of its efficacy, as shown by the K_i value (see Chapter III; *3.9.10*).

References

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K and Watson, J. D. (1994) Molecular Biology of the Cell. Chapter 19. Garland Publishing Inc. ISBN 0-8153-3218-1.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**: 403-410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389-3402.

Arnold, F., West, D. C., Schofield, P. F. and Kumar, S. (1987) Angiogenic activity in human wound fluid. *Int. J. Microcirc. Clin. Exp.* **5**: 381-386

Asai, K., Nakanishi, K., Isobe, I., Eksioglu, Y. Z., Hirano, A., Hama, K., Miyamoto, T. and Kato, T. (1992) Neurotrophic action of gliostatin on cortical neurons. Identity of gliostatin and platelet-derived endothelial cell growth factor. *J. Biol. Chem.* **267**: 20311-20316.

Asai, K., Hirano, T., Kaneko, S., Moriyama, A., Nakanishi, K., Isobe, I., Eksioglou, Y. Z. and Kato, T. A. (1992b) Novel glial growth inhibitory factor, gliostatin, derived from neurofibroma. *J. Neurochem.* **59**: 307-317.

Asai, K., Hirano, T., Matsukawa, K., Kusada, J., Takeuchi, M., Otsuka, T., Matsui, N. and Kato, T. (1993) High concentrations of immunoreactive gliostatin/platelet-derived endothelial cell growth factor in synovial fluid and serum of rheumatoid arthritis. *Clin. Chim. Acta* **218**: 1-4.

Akiyama, S., Furukawa, T., Sumizawa, T., Takebayashi, Y., Nakajima, Y., Shimaoka, S. and Haraguchi, M. (2004) The role of thymidine phosphorylase, an angiogenic enzyme, in tumor progression. *Cancer Sci.* **95**: 851-857.

Balzarini, J., Gamboa, A. E., Esnouf, R., Liekens, S., Neyts, J., De Clercq, E., Camarasa, M. J. and Perez-Perez, M. J. (1998). 7-Deazaxanthine, a novel prototype inhibitor of thymidine phosphorylase. *FEBS Lett.* **438**: 91-95.

Barclay, B. J., Kunz, B. A., Little, J. G. and Haynes, R. H. (1982). Genetic and biochemical consequences of thymidylate stress. *Can. J. Biochem.* **60:** 172-84.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank *Nucleic Acids Res.* **28:** 235-242.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasuni, M. (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur. J. Biochem.* **80**: 319-324.

Birck, M. R. and Schramm, V. L (2004a) Binding causes the remote [5'-^{3}H] Thymidine kinetic isotope effect in human thymidine phosphorylase. *J. Am. Chem. Soc.* **126**: 6882-6883.

Birck, M. R. and Schramm, V. L. (2004b) Nucleophilic participation in the transition state for human thymidine phosphorylase. *J. Am. Chem. Soc.* **126**: 2447-2453.

Black, R. A. and White, J. M. (1998) ADAMS – focus on the protease domain. *Curr. Opin. Cell Biol.* **10**: 654-659.

Blow, D. (2002) Outline of crystallography for biologists. *Oxford University Press.*

Blundell, T. L. and Johnson, L. N. (1976) In protein crystallography. *Academic Press, London.*

Bode, W., Gomis-Ruth, F. X. and Stockler, W. (1993). Astracins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the 'metzincins'. *FEBS Lett.* **331**: 134-140.

Borden, P. and Heller, R. A. (1997) Transcriptional control of matrix metalloproteinases and the tissue inhibitors of matrix metalloproteinases. *Crit. Re. Eukaryot. Gen. Expression* **7**: 159-187.

Bonsor, D., Butz, S. F., Solomons, J., Grant, S., Fairlamb, I. J., Fogg, M. J. and Grogan, G. (2006) Ligation independent cloning (LIC) as a rapid route to families of recombinant biocatalysts from sequenced prokaryotic genomes. *Org. Biomol. Chem.* **4**: 1252-1260.

Bragg, W. H. (1913) The reflection of X-rays by crystals. *Proc. Roy. Soc. London (A)* **88**: 428-438.

Brew, K., Dinakarpandian, D. and Nagase, H. (2000). Tissue inhibitors of metalloproteinases: evolution, structure and function. *Biochim. Biophys. Acta* **1477**: 267-283.

Brodsky, L. I., Vasiliev, A. V., Kalaidzidis, Ya. L., Osipov, Yu. S., Tatuzov, R. L. and Feranchuk, S. I. (1992) GeneBee: the program package for biopolymer structure analysis. *Dimacs* **8**: 127-139.

Brown, N. S. and Bicknell, R. (1998). Thymidine phosphorylase, 2-deoxy-D-ribose and angiogenesis. *Biochem. J.* **334**: 1-8.

Brown, N. S., Jones, A., Fujiyama, C., Harris, A. L. And Bicknell, R. (2000) Thymidine phosphorylase induces carcinoma cell oxidative stress and promotes secretion of angiogenic factors. *Cancer Res.* **60**: 6298-6302.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M.,

Simonson, T. and Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol. Crystallogr. D* **54:** 905-921.

Burling, F. T., Kniewel, R., Buglino, J. A., Chanda, T., Beckwith, A. and Lima, C. D. (2003) Structure of *Escherichia coli* uridine phosphorylase at 2.0Å resolution. *Acta Crystallogr. D.* **59**: 73-76.

Butterwoth, P. J. (1972) The use of Dixon plots to study enzyme inhibition. *Biochim Biophys Acta.* **289**: 251-253.

Caradoc-Davies, T. T., Cutfield, S. M., Lamont, I. L. and Cutfield, J. F. (2004). Crystal structures of *Escherichia coli* uridine phosphorylase in two native and three complexed forms reveal basis of substrate specificity, induced conformational changes and influence of potassium. *J. Mol. Biol.* **337**: 337-354.

Cole, C., Reigan, P., Gbaj, A., Edwards, P. N., Douglas, K. T., Stratford, I. J., Freeman, S. and Jaffar, M. (2003) Potential tumor-selective nitroimidazolylmethyluracil prodrug derivatives: inhibitors of the angiogenic enzyme thymidine phosphorylase. *J. Med. Chem.* **46**: 207-209.

Collaborative Computational Project 4 (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr.* **50**: 760-763.

Collins, J. E., Goward, M. E., Cole, C. G., Smink, L. J., Huckle, E. J., Knowles, S., Bye, J. M., Beare, D. M. and Dunham, I. (2003) Reevaluating human gene annotation: a second-generation analysis of chromosome 22. *Genome Res.* **13**:27-36.

Cheng, J., Randall, A., Sweredoski, M. and Baldi, P. (2005) SCRATCH: a Protein Structure and Structural Feature Prediction Server. *Nucleic Acids Res.* **33**: 72-76.

Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. and Barton, G. J. (1998) Jpred: A Consensus Secondary Structure Prediction Server. *Bioinformatics* **14**: 892-893.

Davies, I. W., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**: 615-619.

Desgranges, C., Razaka, G., Rabaud, M., Bricaud, H., Balzarini, J. and De Clerq, E. (1983) Phosphorolysis of (E)-5-(2-Bromovinyl)-2'-deoxyuridine (BVDU) and other 5-substituted-2'-deoxyuridines by purified human thymidine phosphorylase and intact blood platelets. *Biochem. Pharmacol.* **32**: 3583-3590.

Desgranges, C., Razaka, G., Rabaud, M. and Bricaud, H. (1981) Catabolism of thymidine in human blood platelets. Purification and properties of thymidine phosphorylase. *Biochim. Biophys. Acta* **654**: 211-218.

Dixon, M. (1953) The determination of enzyme inhibitor constants. *Biochem. J.* **55**: 170-171.

Dreusicke, D. and Schulz, G. E. (1986). The glycine-rich loop of adenylate kinase forms a giant anion hole. *FEBS Lett.* **208**: 301-304.

Eda, H., Fujimoto, K., Watanabe, S., Ura, M., Hino, A., Tanaka, Y., Wada, K. and Ishitsuka, H. (1993) Cytokines induce thymidine phosphorylase expression in tumor cells and make them more susceptible to 5'-deoxy-5-fluorouridine. *Cancer Chemother. Pharmacol.* **32**: 333-338.

Edwards, P. N. (2006) A kinetic, modeling and mechanistic re-analysis of thymidine phosphorylase and some related enzymes. *J. Enzyme Inhib. Med. Chem.* **21**:483-499.

Egeblad, M. and Werb, Z. (2002) New functions for the matrix metalloproteinases in cancer progression. *Nat. Rev. Cancer* **2**: 161-174.

El Omari, K., Bronckaers, A., Liekens, S., Perez-Perez, M. J., Balzarini, J. and Stammers, D. K. (2006). Structural basis for non-competitive product inhibition in human thymidine phosphorylase: implications for drug design. *Biochem. J.* **399**: 199-204.

Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **D60**: 2126-2132.

Endo, M., Shinbori, N., Fukase, Y., Sawada, N., Ishikawa, T., Ishitsuka, H. and Tanaka, Y. (1999) Induction of thymidine phosphorylase expression and enhancement of efficacy of capecitabine or 5'-deoxy-5-fluorouridine by cyclophosphamide in mammary tumor models. *Int. J. Cancer* **83**: 127-134.

Ewald, P. P. (1921) *Z. Kristallogra. Miner.* **56**: 129.

Eyrich, V. A. and Rost, B. (2003) META-PP: single interface to crucial prediction servers. *Nucleic Acids Res.* **31**: 3308-3310.

Fernandez-Catalan, C., Bode, W., Huber, R., Turk, D., Calvete, J. J., Lichte, A., Tschesche, H. and Maskos, K. (1998) Crystal structure of the complex formed by the membrane type 1-matrix metalloproteinase with the tissue inhibitor of metalloproteinases-2, the soluble progelatinase A receptor. *EMBO J.* **17**: 5238-5248.

Ferrara, N., Gerber, H. P. and LeCouter (2003) The biology of VEGF and its receptors. *J. Nat. Med.* **9**: 669-676.

Finnis, C., Dodsworth, N., Pollitt, C. E., Carr, G. and Sleep, D. (1993). Thymidine phosphorylase activity of platelet-derived endothelial cell growth factor is responsible for endothelial cell mitogenicity. *Eur. J. Biochem.* **212**: 201-210.

Folkman, J. (1971) Tumour angiogenesis: therapeutic implications. *N. Engl. J. Med.* **285**: 1182-1186.

Folkman, J. (1995) Angiogenesis in cancer, vascular, rheumatoid and other disease. *Nat. Med.* **1**: 27-31.

Folkman, J. (1998) Therapeutic angiogenesis in ischemic limbs. *Circulation* **97** :1108-1110.

Folkman, J. (2002) Role of angiogenesis in tumour growth and metastasis. *Semin. Oncol.* **29**: 15-18.

Folkman, J. (2007) Angiogenesis: an organising principle in drug discovery? *Nature Reviews, Drug Discovery* **6**: 273-286.

Foote, S., Vollrath, D., Hilton, A. and Page, D. C. (1992) The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science.* **258**:60-66.

Friedel, G. (1913). Sur les symétries cristallines que peut révéler la diffraction des rayons X., *C.R. Acad. Sci. Paris*, **157**: 1533-1536.

Friedkin, M. and Roberts, D. (1954) The enzymatic synthesis of nucleoside. I. Thymidine phosphorylase in mammalian tissue. *J. Biol. Chem.* **207**: 245-456.

Fuhrmann, M., Hausherr, A., Ferbitz, L., Schödl, T., Heitzer, M. and Hegemann, P. (2004) Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. *Plant Mol. Biol.* **55**:869-881.

Fukuiwa, T., Takebayashi, Y., Akiba, S., Matsuzaki, T., Hanamure, Y., Miyadera, K., Yamada, Y. and Akiyama, S. (1999) Expression of thymidine phosphorylase and vascular endothelial cell growth factor in human head and neck squamous cell carcinoma and their different characteristics. *Cancer* **85**: 960-969.

Fukushima, M., Suzuki, N., Emura, T., Yano, S., Kazuno, H., Tada, Y., Yamada, Y. and Asao, T. (2000) Structure and activity of specific inhibitors of thymidine phosphorylase to potentiate the function of antitumor 2'-deoxyribonucleosides. *Biochem. Pharmacol.* **59**: 1227-1236.

Fukushima, M., Okabe, H., Takechi, T., Ichikawa, W. and Hirayama, R. (2002) Induction of thymidine phosphorylase by interferon and taxanes occurs only in human cancer cells with low thymidine phosphorylase activity. *Cancer Lett.* **187**: 103-110.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press. pp. 571-607. Copyright- Humana Press.

Gimbrone, M. A., Leapman, S. B., Cotran, R. S., Folkman, J. (1972) Tumour dormancy in vivo by prevention of neovascularisation. *J. Exp. Med.* **136**: 261-276.

Godenschwege, T. A., Pohar, N., Buchner, S. and Buchner, E. (2000). Inflated wings, tissue autolysis and early death in tissue inhibitor of metalloproteinases mutants of *Drosophila*. *Eur. J. Cell Biol.* **79**: 495-501.

Gomis-Rüth, F. X., Maskos, K., Betz, M., Bergner, A., Huber, R., Suzuki, K., Yoshida, N., Nagase, H., Brew, K., Bourenkov, G. P., Bartunik, H., and Bode, W. (1997) Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. *Nature (London)* **389**: 77-81.

Goto, H., Kohno, K., Sone, S., Akiyama, S., Kuwano, M. and Ono, M. (2001) Interferon gamma-dependent induction of thymidine phosphorylase/platelet-derived endothelial growth factor through gamma-activated sequence-like element in human macrophages. *Cancer Res.* **61**: 469-473.

Graeber, T. G., Osmanian, C., Jacks, T., Housman, D. E., Koch, C. J., Lowe, S. W. and Giaccia, A. J. (1996) Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours. *Nature* **379**: 88-91.

Green, D. W., Ingram, V. M. and Perutz, M. F. (1954) The structure of haemoglobin IV. Sign determination by the isomorphous replacement method. *Proc. Roy. Soc.* **A255**: 287-307.

Griffiths, L., Dachs, G. U., Bicknell, R., Harris, A. L. and Stratford, I. J. (1997) The influence of oxygen tension and pH on the expression of platelet-derived endothelial cell growth factor/thymidine phosphorylase in human breast tumor cells grown *in vitro* and *in vivo*. *Cancer Res.* **57**:570-572.

Gross, J. and Lapière, C. M. (1962) Collagenolytic activity in amphibian tissues: a tissue culture assay. *Proc. Natl. Acad. Sci. U S A.* **48**: 1014-1022.

Guex, N. and Peitsch, M. C. (1997) Swiss-model and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**:2714-1723.

Hagiwara, K., Stenman, G., Hona, H., Sahlin, P., Anderson, A., Miyazono, K., Heldin, C. H., Ishikawa, F. and Takaku, F. (1991) Organisation and chromosomal localisation of the human platelet derived endothelial cell growth factor gene. *Mol. Cell Biol.* **11**: 2125-2132.

Hahn, T. (1987) *International tables for Crystallography*. D. Reidel Publishing Company.

Halpert, I., Sires, U. I., Roby, J. D., Potter-Perigo, S., Wight, T. N., Shapiro, S. D., Welgus, H. G., Wickline, S. A. and Parks, W. C. (1996) Matrilysin is expressed by lipid-laden macrophages at sites of potential rupture in atherosclerotic lesions and localizes to areas of versican deposition, a proteoglycan substrate for the enzyme. *Proc. Natl. Acad. Sci. U S A.* **93**:9748-9753.

Hammerberg, C., Fisher, G. J., Voorhees, J. J. and Cooper, K. D. (1991) Elevated thymidine phosphorylase activity in psoriatic lesions accounts for the apparent presence of an epidermal "growth inhibitor," but is not in itself growth inhibitory. *J. Invest. Dermatol.* **97**: 286-290.

Haraguchi, M., Miyadera, K., Uemura, K., Sumizawa, T., Furukawa, T., Yamada, K., Akiyama, S. and Yamada Y. (1994) Angiogenic activity of enzymes. *Nature* **368**:198.

Haraguchi, M., Tsujimoto, H., Fukushima, M., Higuchi, I., Kuribayashi, H., Utsumi, H., Nakayama, A., Hashizume, Y., Hirato, J., Yoshida, H., Hara, H., Hamano, S., Kawaguchi, H., Furukawa, T., Miyazono, K., Ishikawa, F., Toyoshima, H., Kaname, T., Komatsu, M., Chen, Z. S., Gotanda, T., Tachiwada, T., Sumizawa, T., Miyadera, K., Osame, M., Yoshida, H., Noda, T., Yamada, Y. and Akiyama, S. (2002) Targeted deletion of both thymidine phosphorylase and uridine phosphorylase and consequent disorders in mice. *Mol. Cell Biol.* **22**:5212-5221.

Hendrickson, W. A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**: 51-58.

Henriet, P., Blavier, L. and Declerck, Y. A. (1999). Tissue Inhibitors of metalloproteinases (TIMP) in invasion and proliferation. *APMIS* **1**: 111-119.

Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M. C., Wides, R., Salzberg, S. L., Loftus, B. J., Yandell, M. D., Majoros, W. H., Rusch, D. B., Lai, Z., Kraft, C. L., Abril, J. F., Anthouard, V., Arensburger, P., Atkinson, P. W., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chaturverdi, K., Christophides, G. K., Chrystal, M. A. M., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C. A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M. E., Hladun, S. L., Hogan, J. R., Hong, Y. S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C. D., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A. A., Liang, Y., Lin, J. J., Lobo, N. F., Lopez, J. R., Malek, J. A., McIntosh, T. C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S. D., O'Brochta, D. A., Pfannkoch, C., Qi, R., Regier, M. A., Remington, K., Shao, H., Sharakhova, M. V., Sitter, C. D., Shetty, J., Smith, T. J., Strong, R., Sun, J., Thomasova, D., Ton, L. Q., Topalis, P., Tu, Z., Unger, M. F., Walenz, B., Wang, A. H., Wang, J., Wang, M., Wang, X., Woodford, K. J., Wortman, J. R., Wu, M., Yao, A., Zdobnov, E. M., Zhang, H., Zhao, Q., Zhao, S., Zhu, S. C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C. W., Louis, C., Kalush, F., Mural, R. J., Myers, E. W., Adams, M. D., Smith, H. O., Broder, S., Gardner, M. J., Fraser, C. M., Birney, E., Bork, P., Brey, P. T., Venter, J. C., Weissenbach, J., Kafatos, F. C., Collins, F. H. and Hoffman, S. L. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**: 129-149.

Hübscher, U. and Spadari, S. (1994) DNA replication and chemotherapy. *Physiol. Rev.* **74**: 259-304.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti L., De Castro, E., Langendijk-Genevaux, P. S., Pagni, M. and Sigrist, C. J. A. (2006) The Prosite database. *Nucleic Acids Res.* **34**: D227-D230.

Hynes, R. O. and Zhao, Q. (2000). The evolution of cell adhesion. *J. Cell Biol.* **150**: 89-96.

Ieda, Y., Waguri-Nagaya, Y., Iwahasi, T., Otsuka, T., Matsui, N., Namba, M., Asai, K. and Kato, T. (2001) IL-1beta-induced expression of matrix metalloproteinases and gliostatin/platelet-derived endothelial cell growth factor (GLS/PD-ECGF) in a chondrosarcoma cell line (OUMS-27). *Rheumatol. Int.* **21**: 45-52.

Iltzsch, M. H., El Kouni, M. H. and Cha, S. (1985) Kinetic studies of thymidine phosphorylase from mouse liver. *Biochemistry* **24**: 6799-6807.

Imazano, Y., Takebayashi, Y., Nishiyama, K., Akiba, S., Miyadera, K., Yamada, Y., Akiyama, S. and Ohi, Y. (1997) Correlation between thymidine phosphorylase expression and prognosis in human renal cell carcinoma. *J. Clin. Oncol.* **15**: 2570-2578.

Ishihara, M., Obara, K., Ishizuka, T., Fujita, M., Sato, M., Masuoka, K., Saito, Y., Yura, H., Matsui, T., Hattori, H., Kikuchi, M. and Kurita, A. (2003) Controlled release of fibroblast growth factors and heparin from photocrosslinked chitosan hydrogels and subsequent effect on in vivo vascularization *J. Biomed. Mater. Res.* **64A**: 551-559.

Ishikawa, F., Miyazono, K., Hellman, U., Dexler, H., Wernstedt, C., Hagiwara, K., Usuki, K., Takaku, F., Risau, W. and Heldin, C. H. (1989) Identification of angiogenic activity and the cloning and expression of platelet-derived endothelial cell growth factor. *Nature* **338**: 557-562.

Iyer, S., Wei, S., Brew, K. and Acharya, K. R. (2007) Crystal structure of the catalytic domain of matrix metalloproteinase-1 in complex with the inhibitory domain of tissue inhibitor of metalloproteinase-1. *J. Biol. Chem.* **282**: 364-371.

Jones, T. A., Zhou, J. Y., Cowman, S. W. and Kjeldgaard, M. (1991). Improved methods for building models in electron-density maps and the location of errors in these models. *Acta Crystallogr.* **A47**: 110-119.

Kabuubi, P., Loncaster, J. A., Davidson, S. E., Hunter ,R. D., Kobylecki, C., Stratford, I. J. and West, C. M. (2006) No relationship between thymidine phosphorylase (TP, PD-ECGF) expression and hypoxia in carcinoma of the cervix. *Br. J. Cancer* **94**: 115-120.

Karplus, K., Sjölander, K., Barrett, C., Cline, M., Haussler, D., Hughey, R., Holm, L. and Sander, C. (1997) Predicting Protein Structure Using Hidden Markov Models. *Proteins* **1**: 134-139.

Kasugai, K., Joh, T., Kataoka, H., Sasaki, M., Tada, T., Asai, K., Kato, T. and Itoh, M. (1997) Evidence for participation of gliostatin/platelet-derived endothelial cell growth factor in gastric ulcer healing. *Life Sci.* **61**: 1899-1906.

Kashiwagi, M., Tortella, M., Nagase, H. and Brew, K. (2001). TIMP-3 is a potent inhibitor of aggrecanase 1 (ADAM-TS4) and aggrecanase 2 (ADAM-TS5). *J. Biol. Chem.* **276**: 12501-12504.

Kita, T., Takahashi, H. & Hashimoto, Y. (2001). Thymidine phosphorylase inhibitors with a homophthalimide skeleton. *Biol. Pharm. Bull.* **24**: 860-862.

Kitazono, M., Takebayashi, Y., Ishitsuka, K., Takao, S., Tani, A., Furukawa, T., Miyadera, K., Yamada, Y., Aikou, T. and Akiyama, S. (1998) Prevention of hypoxia-induced apoptosis by the angiogenic factor thymidine phosphorylase. *Biochem. Biophys. Res. Commun.* **253**: 797-803.

Kleespies; , A., Bruns, C. J. and Jauch, K. W. (2005) Clinical significance of VEGF-A, -C and -D expression in esophageal malignancies. *Onkologie* **28**:281-288.

Kleywegt, G.J. and Jones, T.A. (1998). Databases in protein crystallography. *Acta Crystallogr.* ***D54***: 1119-1131.

Kline, P. C. and Schramm, V. L. (1993) Purine nucleoside phosphorylase. Catalytic mechanism and transition state analysis of the arsenolysis reaction. *Biochemistry* **32**: 13212-13219.

Kojima, H., Shijubo, N. and Abe, S. (2002) Thymidine phosphorylase and vascular endothelial growth factor in patients with Stage I lung adenocarcinoma. *Cancer* **94**: 1083-1093.

Konttinen, Y. T., Ainola, M., Valleala, H., Ma, J., Ida, H., Mandelin, J., Kinne, R. W., Santavirta, S., Sorsa, T., Lopez-Otin, C. and Takagi, M. (1999) Analysis of 16 different matrix metalloproteinases (MMP-1 to MMP-20) in the synovial membrane: different profiles in trauma and rheumatoid arthritis. *Ann. Rheum. Dis.* **58**: 691-697.

Kornberg, A. and Baker, T. A. (1991) DNA replication; Freeman and Co. New York

Langton, K. P., Barker, M. D. and McKie, N. (1998). Localisation of the functional domains of human tissue inhibitor metalloproteinases-3 and the effects of a Sorby's fundus dystrophy mutation. *J. Biol. Chem.* **273**: 16778-16781.

Lapid, C. (2003) Primer X - Automated design of mutagenic primers for site-directed mutagenesis.

Lakowicz, J. R. (1999) "Principles of Fluorescence Spectroscopy", Plenum Publishing Corporation, 2nd edition.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**: 283-291.

Leatherbarrow, R. J. (2001) GraFit Version 5, Erithacus Software Ltd., Horley, U. K.

Levene, P. A. and Medigreceanu, F. (1911) On nucleases. *J. Biol. Chem.* **9**: 65-83.

Li, J., Brick, P., O`Hare, M. C., Skarzynski, T., Lloyd, L. F., Curry, V. A., Clark, I. M., Bigg, H. F., Hazleman, B. L., Cawston, T. E. and Blow, D. E. (1995) Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed beta-propeller. *Structure* **3**: 541-549.

Liekens, S., Bilsen, F., De Clerq, E., Priego, E. M., Camarasa, M. J., Perez-Perez, M. J. and Balzarini, J. (2002) Antio-angiogenic activity of a novel multi-substrate analogue inhibitor of thymidine phosphorylase. *FEBS Lett.* **510**: 83-88.

Llano, E., Pendas, A. M., Aza-Blanc, P., Kornberg, T. B. and Lopez-Otin, C. (2000) Dm1-MMP, a matrix metalloproteinase from Drosophila with a potential role in extracellular matrix remodeling during neural development. *J. Biol. Chem.* **275**: 35978-35985.

Lovell, S. C., Davis, I. W., Arendall III, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. and Richardson, D. C. (2003) Structure validation by C-alpha geometry: phi, psi, and C-beta deviation. *Proteins* **50**: 437-450.

Lund, L. R., Romer, J., Bugge, T. H., Nielsen, B. S., Frandsen, T.L., Degen, J. L., Stephens, R. W. and Keld Dano (1999) Functional overlap between two classes of matrix-degrading proteases in wound healing. *EMBO J.* **18**: 4645-4656.

Maskos, K., Lang, R., Tschesche, H. and Bode, W. (2007) Flexibility and Variability of TIMP binding: X-ray Structure of the complex between collagenase-3/MMP-13 and TIMP-2. *J. Mol. Biol.* **366**: 1222-1231.

Massova, I., Kotra, L. P., Fridman, R. and Mobashery, S. (1998) Matrix metalloproteinases: structures, evolution and diversification. *FASEB J.* **12**: 1075-1095.

McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. and Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr.* **D61**: 458-464.

McDonald, I. K. and Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**: 777-793.

Mendieta, J., Martin-Santamaria, S., Priego, E. M., Balzarini, J., Camarasa, M. J., Perez-Perez, M. J. and Gago, F. (2004). Role of histidine-85 in the catalytic mechanism of thymidine phosphorylase as assessed by targeted molecular dynamics simulations and quantum mechanical calculations. *Biochemistry* **43**: 405-414.

Michalopoulos, I., Torrance, G. M., Gilbert, D. R. and Westhead, D. R. (2004). TOPS: An enhanced database of protein structural topology. *Nucleic Acid Res.* **32**: D251-D254.

Milkiewicz, M., Ispanovic, E., Doyle, J. L. and Haas, T. L. (2006) Regulators of angiogenesis and strategies for their therapeutic manipulation. *Int. J. Biochem. Cell Biol.* **38**: 333-357.

Miyadera, K., Sumizawa, T., Haraguchi, M., Yoshida, H., Konstanty, W., Yamada, Y. and Akiyama S. (1995) Role of thymidine phosphorylase activity in the angiogenic effect of platelet derived endothelial cell growth factor/thymidine phosphorylase. *Cancer Res.* **55**: 1687-1690.

Miyazono, K., Okabe, T., Urabe, A., Takaku, F. and Heldin, C. H. (1987) Purification and properties of an endothelial cell growth factorfrom human platelets. *J. Biol. Chem.* **262**: 4098-4103.

Miyazono, K., Usiki, K. and Heldin, C. H. (1991) Platelet-derived endothelial cell growth factor. *Progress in Growth Factor Res.* **3**: 207-217.

Miwa, M., Ura, M., Nishida, M., Sawada, N., Ishikawa, T., Mori, K., Shimma, N., Umeda, I. and Ishitsuka, H. (1998). Design of a novel oral fluoropyrimidine carbamate, capecitabine, which generates 5-fluorouracil selectively in tumours by enzymes concentrated in human liver and cancer tissue. *Eur. J. Cancer.* **34**: 1274-1281.

Moghaddam, A., Zhang, H. T., Fan, T. P., Hu, D. E., Lees, V. C., Turley, H., Fox, S. B., Gatter, K. C., Harris A. L. and Bicknell, R. (1995) Thymidine phosphorylase is angiogenic and promotes tumour growth. *Proc. Natl. Acad. Sci. USA* **92**: 998-1002.

Muskett, F. W., Frenkiel, T. A., Feeney, J., Freedman, R. B., Carr, M. D. and Williamson, R. A. (1998) High resolution structure of the N-terminal domain of tissue inhibitor of metalloproteinases-2 and characterisation of its interaction site with matrix metalloproteinase-3. *J. Biol. Chem.* **273**: 21736-21743.

Muro, H., Waguri-Nagaya, Y., Mukofujiwara, Y., Iwahashi, T., Otsuka, T., Matsui, N., Moriyama, A., Asai, K. and Kato, T. (1999) Autocrine induction of gliostatin/platelet-derived endothelial cell growth factor (GLS/PD-ECGF) and GLS-induced expression of matrix metalloproteinases in rheumatoid arthritis synoviocytes. *Rheumatology* **38**: 1195-1202.

Murphy, G. and Gavrilovic, J. (1999) Proteolysis and cell migration: creating a path? *Curr. Opin. Cell Biol.* **11**: 614-621.

Murphy, G. and Knauper, V. (1997). Relating matrix metalloproteinase structure to function: why the 'hemopexin' domain? *Matrix Biol.* **15**: 511-518.

Murphy, G. and Willenbrock, F. (1995). Tissue inhibitors of matrix metalloendopeptidases. *Methods Enzymol.* **248**: 496-510.

Murray, P. E., McNally, V.A., Lockyer, S. D., Williams, K. J., Stratford, I. J., Jaffar, M., and Freeman, S. (2002) Synthesis and enzymatic evaluation of pyridinium-substituted uracil derivatives as novel inhibitors of thymidine phosphorylase. *Bioorganic & Medicinal Chem.* **10**:525-530.

Murshudov, G. N., Vagin, A. A. and Dodson, E. D. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **D53**: 1285-1294.

Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**: 536-540.

Nagase, H. and Woessner, J. F. (1999) Matrix metalloproteinases. *J. Biol. Chem.* **274**: 21491-21494.

Nagase, H. (1998). Cell surface activation of progelatinase A (proMMP-2) and cell migration. *Cell Res.* **8:** 179-186.

Nagase, H. and Brew, K. (2002) Engineering of tissue inhibitor of metalloproteinases mutants as potential therapeutics. *Arthritis Res.* **4**: S51-S61.

Nakajima, Y., Gotanda, T., Uchimiya, H., Furukawa, T., Haraguchi, M., Ikeda, R., Sumizawa, T., Yoshida, H. and Akiyama, S. (2004) Inhibition of metastasis of tumor cells overexpressing thymidine phosphorylase by 2-deoxy-L-ribose. *Cancer Res* **64**: 1794-1801.

Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr.* **A50**: 157-163.

Nelson, A. R., Fingleton, B., Rothenberg, M. L. and Matrisian, L. M. (2000) Matrix metalloproteinases: biologic activity and clinical implications. *J. Clin. Oncol.* 18: 1135-1149.

Nene, V., Wortman, J. R., Lawson, D., Haas, B., Kodira, C. D., Tu, Z. J., Loftus, B., Xi, Z., Megy, K., Grabherr, M., Ren, Q., Zdobnov, E. M., Lobo, N. F., Campbell, K. S., Brown, S. E., Bonaldo, M. F., Zhu, J., Sinkins, S. P., Hogenkamp, D. G., Amedo, P., Arsenburger, P., Atkinson, P. W., Bidwell, S., Biedler, J., Birney, E., Bruggner, R. V., Costas, J., Coy, M. R., Crabtree, J., Crawford, M., DeBruyn, B., DeCaprio, D., Eiglmeier, K., Eisenstadt, E., El-Dorry, H., Gelbart, W. M., Gomes, S. L., Hammond, M., Hannick, L. I., Hogan, J. R., Holmes, M. H., Jaffe, D., Johnston, S. J., Kennedy, R. C., Koo, H., Kravitz, S., Kriventseva, E. V., Kulp, D., Labutti, K., Lee, E., Li, S., Lovin, D. D., Mao, C., Mauceli, E., Menck, C. F., Miller, J. R., Montgomery, P., Mori, A., Nascimento, A. L., Naveira, H. F., Nusbaum, C., O'leary, S. B., Orvis, J., Pertea, M., Quesneville, H., Reidenbach, K. R., Rogers, Y. H., Roth, C. W., Schneider, J. R., Schatz, M., Shumway, M., Stanke, M., Stinson, E. O., Tubio, J. M. C., Vanzee, J. P., Verjovski-Almeida, S., Werner, D., White, O. R., Wyder, S., Zeng, Q., Zhao, Q., Zhao, Y., Hill, C. A., Raikhel, A. S., Soares, M. B., Knudson, D. L., Lee, N. H., Galagan, J., Salzberg, S. L., Paulsen, I. T., Dimopoulos, G., Collins, F. H., Bruce, B., Fraser-Liggett, C. M. and Severson, D. W. (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* **316**: 1718-1723.

Norman, R. A., Barry, S. T., Bate, M., Breed, J., Colls, J. G., Ernill, R. J., Luke, R. W., Minshull, C. A., McAlister, M. S., McCall, E. J., McMiken, H. H., Paterson, D. S., Timms, D., Tucker, J. A. and Pauptit, R. A. (2004). Crystal structure of human thymidine phosphorylase in complex with a small molecule inhibitor. *Structure* **12:** 75-84.

Norrby, K. (1997) Angiogenesis: new aspects relating to its initiation and control. *APMIS* **105**: 417-437.

Notredame, C., Higgins, D. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**: 205-217.

O'Brien, R., Ladbury, J. E. and Chowdry B. Z. (2000) Isothermal titration calorimetry of biomolecules. Chapter 10 in Protein-Ligand interactions: hydrodynamics and calorimetry Ed. Harding, S.E. and Chowdry, B.Z, Oxford University Press.

O'Byrne, K. J., Koukourakis, M. I., Giatromanolaki, A., Cox, G., Turley, H., Steward, W. P., Gatter, K. and Harris, A. L. (2000) Vascular endothelial growth factor, platelet-derived endothelial cell growth factor and angiogenesis in non-small-cell lung cancer. *Br. J. Cancer* **82**: 1427-1432.

Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. and Thornton, J. M. (1997) CATH- A Hierarchic Classification of Protein Domain Structures. *Structure* **5**: 1093-1108.

Otwinowski, Z. (1993a) DENZO: an oscillation data processing programme for macromolecular crystallography. Yale University, New Haven, CT.

Otwinowski, Z. (1993b) SCALEPACK: Software for the scaling together of integrated intensities measured on a number of separate diffraction images. Yale University, New Haven, CT.

Paege, L. M. and Schlenk, F. (1952) Bacterial uracil ribose phosphorylase. *Arch. Biochem. Biophys.* **40**: 42-49.

Park, H. I., Ni, J., Gerkema, F. E., Liu, D., Belozerov, V. E. and Sang, Q. X. (2000) Identification and characterisation of human endometase (Matrix metalloproteinase-26) from endometrial tumor. *J. Biol. Chem.* **275**: 20540-20544.

Parks, W. C. and Mecham, R. P. (1998). *Matrix Metalloproteinases* (Press, A., Ed.), San Diego.

Pearson, W. R. (1990) Rapid and Sensitive sequence comparison with FASTP and FASTA. *Methods in Enzymol.* **183**: 63-98.

Pearson, W. R. and Lipman, D. J. (1998) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* **85**: 2444-2448.

Piper, A. A. and Fox, R. M. (1982) Biochemical basis for the differential sensitivity of human T- and B-Lymphocyte lines to 5-Fluorouracil. *Cancer Res.* **42**: 3753-3760.

Pohar, N., Godenschwege, T. A. and Buchner, E. (1999). Invertebrate tissue inhibitor of metalloproteinase: structure and nested gene organisation within the synapsin locus is conserved from *drosophila* to human. *Genomics* **57:** 293-296.

Poirot, O., O'Toole, E. and Notredame, C. (2003) Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. *Nucleic Acids Res.* **31**: 3503-3506.

Puente, X. S., Sanchez, L. M., Overall, C. M. and Lopez-Otin, C. (2003) Human and mouse proteases: a comparative genomic approach. *Nat. Rev. Genet.* **4**: 544-558.

Pugmire, M. J. and Ealick, S. E. (1998) The crystal structure of pyrimidine nucleoside phosphorylase in a closed conformation. *Structure* **6**: 1467-1479.

Pugmire, M. J., Cook, W. J., Jasanoff, A. Walter, M. R. and Ealick, S. E. (1998a) Structural and theoretical studies suggest domain movement produces an active conformation of thymidine phosphorylase. *J. Mol. Biol.* **281**: 285-299.

Pugmire, M. J. and Ealick S. E. (2002) Structural analyses reveal two distinct families of nucleoside phosphorylases. *Biochem. J.* **361**: 1-25.

Pugmire, M. J., Cook, W. J., Jasanoff, A., Walter, M. R. and Ealick, S. E. (1998) structural and theoretical studies suggest domain movement produces an active conformation of thymidine phosphorylase. *J. Mol. Biol.* **281**: 285-299.

Ramachandran, G., Ramakrishnan, C. and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**: 95-99.

Rawlings, N. D., Morton, F. R. and Barrett, A. J. (2006) *MEROPS*: the peptidase database. *Nucleic Acids Res.* **34**: 270-272.

Razzell, W. E. and Casshyap, P. (1964) Substrate specificity and induction of thymidine phosphorylase in Escherichia coli. *J. Biol. Chem.* **239**: 1789-1793.

Read, R. J. (2001) Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr.* **D57**: 1373-1382.

Reigan, P., Edwards, P. N., Gbaj, A., Cole, C., Barry, S. T., Page, K. M., Ashton, S. E., Luke, R. W., Douglas, K. T., Stratford, I. J., Jaffar, M., Bryce, R. A. and Freeman, S. (2005) Aminoimidazolylmethyluracil analogues as potent inhibitors of thymidine phosphorylase and their bioreductive nitroimidazolyl prodrugs. *J. Med. Chem.* **48**: 392-402.

Reynolds, K., Farzaneh, F., Collins, W. P., Campbell, S., Bourne, T. H., Lawton, F., Moghaddam, A., Harris, A.L. and Bicknell R. (1994) Association of ovarian malignancy with expression of platelet-derived endothelial cell growth factor. *J. Natl. Cancer Inst.* **16**: 1234-1238.

Rick, S. W., Abashkin, Y. G., Hilderbrandt, R. L. and Burt, S. K. (1999). Computational studies of the domain movement and the catalytic mechanism of thymidine phosphorylase. *Proteins* **37**: 242-252.

Rossmann, M. and Blow, D. (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystalogr.* **15**: 24-31.

Rost, B., Sander, C. and Schneider, R. (1994) PHD – an automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.* **10**: 53-60.

Saini, H. K. and Fischer, D. (2005) Meta-DP: domain prediction meta-server. *Bioinformatics* **21**: 2917-2920.

Saito, S., Tsuno, N. H., Sunami, E., Hori, N., Kitayama, J., Kazama, S., Okaji, Y., Kawai, K., Kanazawa, T., Watanabe, T., Shibata, Y. and Nagawa, H. (2003) Expression of platelet-derived endothelial cell growth factor in inflammatory bowel disease. *J. Gastroenterol.* **38**: 229-237.

Sambrook, J., Fritsch, E. F. and Maniatis, T., Eds. (1989). Molecular Cloning: a laboratory manual. Second edition. Edited by Press, C. S. H. L., New York: Cold Spring Harbor.

Sawada, N., Ishikawa, T., Fukase, Y., Nishida, M., Yoshikubo, T.and Ishitsuka, H. (1998) Induction of thymidine phosphorylase activity and enhancement of capecitabine efficacy by taxol/taxotere in human cancer xenografts. *Clin. Cancer Res.* **4**: 1013-1019.

Schwede, T., Kopp, J., Guex, N. and Peitsch, M. C. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **31**: 3381-3385.

Seki, N., Kodama, J., Hongo, A., Miyagi, Y., Yoshinouchi, M. and Kudo, T. (1999) Angiogenesis and platelet-derived endothelial cell growth factor/thymidine phosphorylase expression in endometrial cancer. *Int. J. Oncol.* **15**: 781-786.

Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci U S A.* **89**:8794-8797.

Shweiki, D., Itin, A., Soffer, D. and Keshet, E. (1992) Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis. *Nature* **359**: 843-845.

Sivridis, E., Giatromanolaki, A., Papadopoulos, I., Gatter, K. C., Harris, A. L and Koukourakis, M. I. (2002) Thymidine phosphorylase expression in normal, hyperplastic and neoplastic prostates: correlation with tumour associated macrophages, infiltrating lymphocytes, and angiogenesis. *Br. J. Cancer* **86**: 1465-1471.

Soncin, F., Strydom, D. J. and Shapiro, R. (1997) Interaction of heparin with human angiogenin. *J. Biol. Chem.* **272**:9818-9824.

Spadari, S., Giarrocchi, G., Focher, F., Verri, A., Maga, G., Arcamone, F., Iafrate, E., Manzini, S., Garbesi, A. and Tondelli, L. (1995) 5-iodo-2'-deoxy-L-uridine (L-BVDU): selective phosphorylastion by herpres simplex virus type I thymidine kinase, antiherpetic activity and cytotoxicity studies. *Mol. Pharmacol.* **47**: 1231-1238.

Spinazzola, A., Marti, R., Nishino, I., Andreu, A. L., Naini, A., Tadesse, S., Pela, I., Zammarchi, E., Donati, M. A., Oliver, J. A. and Hirano, M. (2002) Altered thymidine metabolism due to defects of thymidine phosphorylase. *J. Biol. Chem.* **277**: 4128-4133.

Spraggon, G., Stuart, D., Ponting, C., Finnis, C., Sleep, D. and Jones, Y. (1993). Crystallisation and x-ray diffraction study of recombinant platelet-derived endothelial cell growth factor. *J. Mol. Biol.* **234**: 879-880.

Sternlicht, M. D. and Werb, Z. (2001). How matrix metalloproteinases regulate cell behaviour. *Annu. Rev. Cell Dev. Biol.* **17**: 463-515.

Storoni, L. C., McCoy, A. J. and Read, R. J. (2004). Likelihood-enhanced fast rotation functions. *Acta Crystallogr.* ***D*60**: 432-438.

Takao S., Takebayashi, Y., Che, X., Shinchi, H., Natsugoe, S., Miyadera, K., Yamada, Y., Akiyama, S. and Aikou, T. (1998) Expression of thymidine phosphorylase is associated with a poor prognosis in patients with ductal adenocarcinoma of the pancreas. *Clin. Cancer Res.* **4**: 1619-1624.

Takebayashi, Y., Akiyama, S., Akiba, S., Yamada, K., Miyadera, K., Sumizawa, T., Yamada, Y., Murata, F. and Aikou, T. (1996) Clinicopathologic and prognostic significance of an angiogenic factor, thymidine phosphorylase, in human colorectal carcinoma. *J. Natl. Cancer Inst.* **88**: 1110-1117.

Takeuchi, M., Otsuka, T., Matsui, N., Asai, K., Hirano, T., Moriyama, A., Isobe, I., Eksioglu, Y. Z., Matsukawa, K., Kato, T. and Toyohiro, T. (1994) Aberrant production of gliostatin/platelet-derived endothelial cell growth factor in rheumatoid synovium. *Arthritis Rheum.* **37**: 662-672.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-4680.

Toi, M., Inada, K., Hoshina, S., Suzuki, H., Kondo, S. and Tominaga, T. (1995a) Vascular endothelial growth factor and platelet-derived endothelial cell growth factor are frequently co-expressed in highly vascularised human breast cancer. *Clin. Cancer Res.* **1**: 961-964.

Toi, M., Ueno, T., Matsumoto, H. et al. (1995b) Significance of thymidine phosphorylase as a marker of protumour monocytes in breast cancer. *Clin. Cancer Res.* **5**:1131-1137.

Toi, M., Rahman, M. A., Bando, H. and Chow, L. W. C. (2005) Thymiodine phosphorylase (platelet-derived endothelial-cell growth factor) in cancer biology and treatment. *Lancet Oncol.* **6**: 158-166.

Tronrud, D. E. (1997). The TNT refinement package in macromolecular crystallography. *Methods in Enzymol.* **277**: 306-319.

Tuuttila, A., Morgunova, E., Bergmann, U., Lindqvist, Y., Maskos, K., Fernandez-Catalan, C., Bode, W., Tryggvason, K. and Schneider, G. (1998) Three-dimensional structure of human tissue inhibitor of metalloproteinases-2 at 2.1Å resolution. *J. Mol. Biol.* **284**: 1133-1140.

Uría, J. A. and López-Otín, C. (2000) Matrilysin-2, a new matrix metalloproteinase expressed in human tumors and showing the minimal domain organization required for secretion, latency, and activity. *Cancer Res.* **60**: 4745-4751.

Usuki, K., Heldin, N. H., Miyazono, K., Ishikawa, F., Takaku, F., Westernmark, B. and Heldin, C. H. (1989) Production of platelet-derived endothelial cell growth factor by normal and transformed cell in culture. *Proc. Natl. Acad. Sci. USA* **86**: 7427-7431.

Usuki, K., Saras, J., Waltenberger, J., Miyazono, K., Pierce, G., Thomason, A. and Heldin, C. H. (1992) Platelet-derived endothelial cell growth factor has thymidine phosphorylase activity. *Biochem. Biophys. Res. Commun.* **184**: 1311-1316.

Vagin, A. and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**: 1022-1025.

Van Wart, H.E. and Birkedal-Hansen, H. (1990) The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. *Proc. Natl. Acad. Sci. U S A.* **87**: 5578-5582.

Verri, A., Focher, F., Duncomber, R. J., Basnak, I., Walker, R. T., Coe, P. L., De Clerq, E., Andrei, G., Snoeck, R., Balzarini, J. and Spadari, S. (2000) Anti-(herpes vimplex virus) activity of 4'-thio-2'deoxyuridines: a biological investigation for viral and cellular target-enzymes. *Biochem. J.* **351**: 319-326.

Vita, A., Amici, A., Cacciamani, T., Lancotti, M. and Magni, G. (1986) Uridine phosphorylase from Escherichia coli B. Enzymatic and molecular properties. *Int. J. Biochem.* **18**: 431-436.

Vogel, T., Blake, D. A., Whikehart, D. R., Guo, N. H., Zabrenetzky, V. S. and Roberts, D. D. (1993) Specific simple sugars promote chemotaxis and chemokinesis of corneal endothelial cells. *J. Cell Physiol.* **157**: 359-366.

Vu, T. and Werb, Z. (2000) Matrix metalloproteinases: effectors of development and normal physiology. *Genes Dev.* **14**: 2123-2133.

Waguri, Y., Otsuka, T., Sugimura, I., Matsui, N., Asai, K., Moriyama, A. and Kato, T. (1997) Gliostatin/platelet-derived endothelial cell growth factor as a clinical marker of rheumatoid arthritis and its regulation in fibroblast-like synoviocytes. *Br. J. Rheumatol.* **36**: 315-321.

Walter, M. R., Cook, W. J., Cole, L. B., Short, S. A., Koszalka, G. W., Kenitsky, T. A. and Ealick, S. E. (1990) Three-dimensional structure of thymidine phosphorylase from *Escherichia coli* at 2.8Å resolution. *J. Biol. Chem.* **265**: 14016-14022.

Wei, S., Xie, Z., Filenova, E. and Brew, K. (2003). *Drosophila* TIMP is a potent inhibitor of MMPs and TACE: similarities in structure and function to TIMP-3. *Biochemistry* **42**: 12200-12207.

Werb, Z. (1997) ECM and cell surface proteolysis: regulating cellular ecology. *Cell* **91**: 439-442.

Williamson, R. A., Marston, F. A., Angal, S., Koklitis, P., Panico, M., Morris, H. R., Carne, A. F., Smith, B. J., Harris, T. J. and Freedman, R. B. (1990). Disulphide bond assignment in human tissue inhibitor of metalloproteinases (TIMP). *Biochem. J.* **268**: 267-274.

Williamson, R. A., Martorell, G., Carr, M. D., Murphy, G., Docherty, A. J., Freedman, R. B. and Feeney, J. (1994) Solution structure of the active domain of tissue inhibitor of metalloproteinases-2. A new member of the OB fold protein family. *Biochemistry.* **33**: 11745-11759.

Winn, M., Isupov, M. and Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr.* **D57**: 122-133.

Yao, K. S., Clayton, M. and O'Dwyer, P. J. (1995) Apoptosis in human adenocarcinoma HT29 cells induced by exposure to hypoxia. *J. Natl. Cancer Inst.* **87**: 117-122.

Yao, Y., Kubota, T., Sato, K. and Kitai, R. (2001) Macrophage infiltration-associated thymidine phosphorylase expression correlates with increased microvessel density and poor prognosis in astrocytic tumors. *Clin. Cancer Res.* **7**: 2041-2046.

Yoshimura, A., Kuwazuru, Y., Furukawa, T., Yoshida, H., Yamada, K. and Akiyama, S. (1990). Purification and tissue distribution of human thymidine phosphorylase; high expression in lymphocytes, reticulocytes and tumours. *Biochim. Biophys. Acta*, **1034**: 107-113.

Yu, Q. and Stamenkovic, I. (2000) Cell surface-localized matrix metalloproteinase-9 proteolytically activates TGF-beta and promotes tumor invasion and angiogenesis. *Genes Dev.* **14**: 163-176.

Zhang, H. T., Craft, P., Scott, P. A., Ziche, M., Weich, H. A., Harris, A. L. and Bicknell, R. (1995) Enhancement of tumor growth and vascular density by transfection of vascular endothelial cell growth factor into MCF-7 human breast carcinoma cells. *J. Natl. Cancer Inst.* **87**: 213-219.

http://merops.sanger.ac.uk/

http://www.smart.embl-heidelberg.de/smart/do_annotation.pI?BLAST=DUMMY&ACC=SM00206.

http://au.expasy.org/cgi-bin/sprot-search-ful?makeWild=&SEARCH=TIMP.

http://www.ncbi.nlm.nih.gov/BLAST/.

http://www.ysbl.york.ac.uk/~tom/structure.html#2D-prediction